THE APPLICATION OF KRIGING
IN THE STATISTICAL ANALYSIS
OF ANTHROPOMETRIC DATA
VOLUME I

THESIS

Michael Grant
Captain, USAF

AFIT/GOR/ENY/ENS/90M-8

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY
AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT/GOR/ENY/ENS/90M-8

THE APPLICATION OF KRIGING IN THE STATISTICAL ANALYSIS OF ANTHROPOMETRIC DATA VOLUME I

THESIS

Michael Grant
Captain, USAF

AFIT/GOR/ENY/ENS/90M-8

Approved for public release; distribution unlimited

AFIT/GOR/ENY/ENS/90M-8

THE APPLICATION OF KRIGING IN THE STATISTICAL ANALYSIS OF ANTHROPOMETRIC DATA VOLUME I

THESIS

Presented to the Faculty of the School of Engineering
of the Air Force Institute of Technology
Air University
In Partial Fulfillment of the
Requirements for the Degree of
Master of Science in Operations Research

Michael Grant, B.S., M.S.
Captain, USAF

March 1990

Approved for public release; distribution unlimited

Preface

This thesis develops the application of kriging in the statistical analysis of anthropometric data to support improvements in the design of flight equipment. As a result of this research, engineers now have the statistical data to support the design of flight apparatus which accounts for the shape of facial features and the variances between individuals. The benefits of this study will include the enhanced flight performance of crew members realized from the improved comfortability and functional precision of newly designed flight apparatus and the reduction in custom fittings of currently used equipment which do not accurately account for shape or variability. In addition to the cost savings and increased mission capability, this research may potentially benefit developers of medical equipment such as operating room oxygen masks and prosthetic devices.

Special recognition is given to my advisor, Major David G. Robinson, who conceived and developed the idea of adapting the estimation technique of kriging from the field of geostatistics and applying it to the field of anthropometry. In addition to his insights, I offer my personal thanks for his guidance and assistance in completing this research effort.

I would also like to thank the other members of my committee, Major Kenneth W. Bauer and Lt Col James N. Robinson. As co-advisor, Major Bauer provided invaluable guidance and support, particularly in the multivariate analysis. Lt Col Robinson helped immensely in the learning of IDL and offered his support throughout the research effort.

This study was sponsored by the Human Engineering Division of the Harry G. Armstrong Aerospace Medical Research Laboratory. I would like to thank the laboratory and the contractor personnel who assisted in this effort. In particular, I wish to thank Kathleen Robinette(AAMRL/HEG) for enlightening me on the field

of anthropometrics and the challenges of designing flight equipment to the shape and variability of human faces. Also, I would like to thank Greg Zehner(AAMRL/HEG), Joyce C. Robinson(Scientific Research Laboratories) and Bob Beecher(Beecher Associates) for their help with the data collection and manipulation.

Most of all, I thank my wife, Audrey, and sons, Christopher and Matthew, for their support, patience, and understanding throughout the course of this effort. They are my inspiration.

Michael Grant

Table of Contents

List of Figures

List of Tables

AFIT/GOR/ENY/ENS/90M-8

Abstract

Quality flight equipment is essential to flight crew safety and performance. Oxygen masks, night-vision goggles, and other apparatus must fit crew members comfortably and with complete functional precision. A problem currently facing the Air Force is the inconsistent quality of flight equipment. As new equipment is developed to improve crew members' performance, the requirement for design engineers to accurately account for the shape and variability of facial features becomes more critical.

This thesis develops the application of kriging in the statistical analysis of anthropometric data to support improvements in the design of flight equipment. Specifically, the geostatistical estimation technique of kriging is used to estimate the facial surfaces which influence the designs of flight apparatus. These surfaces account for the shape of the facial features and minimize the variance between individuals. A Kalman filter is developed to update and aggregate the kriged surfaces. As a proof of concept study, the techniques are demontrated using data to support the design of the night-vision goggles currently under development. To further enhance the surface estimates, a multivariate analysis is performed to identify the parameters which account for the majority of the variability between faces and to group the faces into homogenous clusters.

THE APPLICATION OF KRIGING IN THE STATISTICAL ANALYSIS OF ANTHROPOMETRIC DATA

I. Introduction

This research effort develops the application of kriging to the statistical analysis of anthropometric data in support of improvements in the design of flight equipment. Kriging is a geostatistical estimation procedure named after D.G. Krige, a South African mining engineer. This study was sponsored by the Human Engineering Division of the Harry G. Armstrong Aerospace Medical Research Laboratory (AAMRL). This first chapter provides the background, the research objectives, and the scope of the study.

Background

Quality flight equipment is essential to flight crew safety and performance. Oxygen masks, night-vision goggles, helmets, and other apparatus must fit crew members comfortably and with complete functional precision. A problem currently facing the Air Force is the inconsistent quality of flight equipment. The poor fit of existing oxygen masks and the requirement that the new night-vision goggles be designed such that one size will fit the entire population of crew members are only two examples which illustrate the need for improvements in flight equipment designs.

According to Kathleen Robinette, a research physical anthropologist in the Workload and Ergonometrics Branch at AAMRL, an estimated 30 to 40 oxygen masks per month must be custom fitted for crew members because the current mask, designated MBU 12/P, is not suited to the variability of facial features among individuals (20). More specifically, the shape of the mask is not adequately specified and

this results in the passing of air through the contact surfaces and into the eyes of the crew members, movement of the mask on the face during high G maneuvers, and a general dissatisfaction with comfort. To further complicate the problem, the custom constructions can only support the fitting of a previous mask design, the MBU 5/P, which does not function as well as the MBU 12/P in a high G environment (20).

The problem of designing flight equipment to account for the shape of human facial features continues as new systems are being developed. The Eagle Eye night-vision goggles are being developed by the Night Vision Corporation on the supposition that one size will fit all crew members. To ensure proper fit, the designers will need to consider the shape in the area of the eyes and nose as well as the length and width of the facial region. As more sophisticated systems are being developed, the need for more precise sizing procedures increases. One such system is the High Altitude Low Pressure System (HALPS). This system consists of a complete flight ensemble which includes an oxygen mask for forcing oxygen into the crew member's respiratory system to improve performance in high G maneuvers. Traditionally, caliper measurements are used in the sizing procedures and these measurements do not accurately account for the spatial variation of the objects under study (20). To improve the quality of flight equipment, alternative statistical methods for representing anthropometric data must be considered.

Researchers at AAMRL have supported the development of a technique for measuring three dimensional surfaces using lasers. The new measurement technique provides contour points which can be analyzed to provide designers with information on facial regions. The laser scans are made using a 4020/PS Echo Digitizer (manufactured by Cyberware Inc.) which measures the distance from the surface being scanned to a point on the arm which holds the laser as the arm rotates around the subject. As a result of this new method, distances between points can be calculated without using calipers and more information on the shape of the facial surface is available.

Having solved the problem of data collection, researchers are now concerned with how to analyze the data to provide the engineers with properly specified design requirements. The goal is to find a representative surface of a particular facial region which will account for the shape of the region and the variablity of the facial features between individuals. If the data can be represented in this manner, the design engineers will be able to develop flight equipment which will fit more comfortably and with better functional precision.

Research Objectives

The purpose of this research effort was to statistically analyze anthropometric data to support improvements in the design of flight equipment. The goal was to develop a procedure for estimating the surface region in the area where the flight apparatus is worn which would minimize the variability between individuals and account for the shape of the region. The resulting estimates would be used by design engineers in the development of new flight equipment. Specifically, the following objectives were established.

Procedure Development. The first objective was to develop a viable kriging procedure for estimating the facial surfaces. Analogies between facial and geological surfaces merited an investigation into determining the feasibility of applying the geostatistical estimation technique of kriging in the estimation of anthropometric surfaces. The development of this procedure was based on the theory of kriging from the field of geostatistics and was adapted to the peculiarites of the anthropometric data.

Aggregation of Individual Estimates. The second objective of this study was to develop a recursive model for aggregating and updating the individual surface estimates. The recursive model was required to facilitate the revision of the surface estimates and to minimize the amount of data required to update the estimates as

more data became available.

Facial Surface Estimation. A third objective was to apply the kriging procedure to estimate the facial region affected by the contact surfaces of the night-vision goggles currently under design. As a proof of concept study, this application demonstrated the feasibility of the technique and provides the foundation for further study in this area. The kriging procedure developed in this research effort was not unique to the estimation of the surface region for the night-vision goggles and could be applied to other anthropometric surfaces.

Facial Classification. A final objective of the study was to determine if the faces could be grouped according to specific parameters prior to kriging and updating. The hypothesis was that these groups would correspond to potential sizes for the flight apparatus and that the surface estimates would be improved further if the sizes were estimated independently. Because only one size was planned for the night-vision goggles, this analysis was performed independent of the goggles study. However, the primary focus of this research was in the statistical analysis of anthropometric data and the estimation of surface regions which would minimize the variablity between facial features. Therefore, the potential for clustering faces based on minimal dimensionality was explored.

Scope

This thesis develops and demonstrates the application of kriging in the analysis of anthropometric data. Specifically, this study includes a discussion of the theoretical development of kriging, the computer programs necessary for the application of the techniques, the documentation of the night-vision goggles application, and a multivariate analysis of facial dimensions to determine potential subgroupings of the subjects. The following provides a summary of the extent of this research effort.

Theory of Kriging. The literature review provides an introduction to the theory and the development of kriging from the field of geostatistics. Emphasis is placed on the the fundamental kriging equations and structural analysis of the data.

Data Collection and Orientation. A description of the data collection, manipulation, and orientation processses are provided. Additionally, the computer programs used in reformatting the data, graphically displaying the data, and aligning the facial regions for each subject on a common coordinate system are provided.

Procedural Development. A complete development of the kriging procedure is outlined in the Chapter III.

Kriging Programs. This document includes a complete package of the programs required in the kriging of anthropometric data. Specifically, the following programs are included.

Experimental Variogram Program. This program determines the experimental variograms for each subject. The program is written in FORTRAN.

Least Squares Variogram Estimation Program. This program is written in FORTRAN and determines the least squares parameters for three of the more commonly used variogram models: the linear, the De Wijsian, and the spherical models.

Kriging Program. This program estimates the values of the facial surface throughout the range of design points and provides the kriging variances for these points. The program is written in C.

Recursive Model. The development of a recursive model for updating the overall surface estimates and variances is provided. The literature review includes a brief introduction to Bayesian statistics and Kalman Filtering. Additionally, this document includes the graphical representations of the updated surfaces from the night-vision goggles study.

Surface Estimates. A description of the process used in obtaining the facial estimates for the night-vision goggles and the graphical representations of the estimates are provided.

Multivariate Analysis. This research includes the factor analysis of various distance and angular measures to determine a new set of measures which would account for the common variations expressed in the original set of variables. Furthermore, the technique of clustering was used to construct homogenous groups within the the subjects based on the variabies identified in the factor analysis. The literature review introduces the multivariate techniques. The data extraction programs and the SAS procedures are also provided.

II. Literature Review

This chapter provides a review of the literature pertaining to the areas of kriging, structural analysis, Bayesian statistics, and multivariate analysis to the degree that they are developed in this thesis. The emphasis of this review is on the origin and development of kriging in the field of geostatistics and the structural analysis essential to the application of the kriging procedures.

Kriging

A complete review of the literature shows no documented applications of kriging in the analysis of anthropometric data. Therefore, this review considers the theory of the technique as developed in the field of geostatistics. Specifically, the following kriging topics are discussed: the origin, a definition, the fundamental equations, the assumptions, and several types of kriging.

Origin of Kriging. The technique of kriging is considered by Clark to be "the simplest application of the Theory of Regionalised Variables" (3, 1). Georges Matheron is credited with introducing the concept of regionalized variables and Clark states that "the application of this theory to problems in geology and mining has lead to the more popular name Geostatistics" (3, 1). Davis explains that geostatistics was devised by Matheron "to treat problems that arise when conventional statistical theory is used in estimating changes in ore grade within a mine" (6, 239). According to Matheron:

> Geostatistics, in their most general acceptation, are concerned with the study of the distribution in space of useful values for mining engineers and geologists, such as grade, thickness, or accumulation, including a most important practical application of the problems arising in ore-deposit evaluation (14, 224).

Journel states that "in mining practice, one problem is to find the best possible estimator of the mean grade of a block" (12, 563) He further states that D.G. Krige proposed a regression technique for this problem in 1951 and that "in 1963, Matheron formalized and generalized this regression procedure and gave it the name of kriging" (12, 563) As David suggests, "the particular nature of estimation problems in mine planning is such that it most probably deserves the use of a special name" (5, 237). At the time, D.G. Krige was a mining engineer in South Africa (6, 383).

Definition. According to the original definition given by Matheron, "kriging is the probabilistic process of obtaining the best linear unbiased estimator of an unknown variable" (12, 563). Stated another way, "kriging is a local estimation technique which provides the best linear unbiased estimator (abbreviated BLUE) of the unknown characteristic studied" (13, 304). In this context, "best" is defined as "having the smallest estimation variance" (3, 104). Matheron later generalized techniques for obtaining nonlinear unbiased estimates and used the name kriging in describing them because they also minimized the estimation variance (12, 563–564). Because of the varied use of the name kriging, Journel suggests that kriging should be redefined as "a probabilistic theory of estimation based on the principle of minimization of the estimation variance" (12, 563).

In geostatistics, the estimation process is typically directed toward determination of the value of an ore deposit at an unsampled location. Clark provides an example of estimating both the value of an uranium deposit at a specific location and the average value of uranium over an area or block (3, 69– 74). Davis introduces an example of determining the water table elevations at unknown points based on the values known at different well sites (6, 386–392). To clarify the definition of kriging and to illustrate the potential uses of this technique, the following example adapted from *Developments in Geomathematics* is provided (1, 353).

Consider the five control points labeled $S_1 - S_5$ in Figure 2.1. These points are

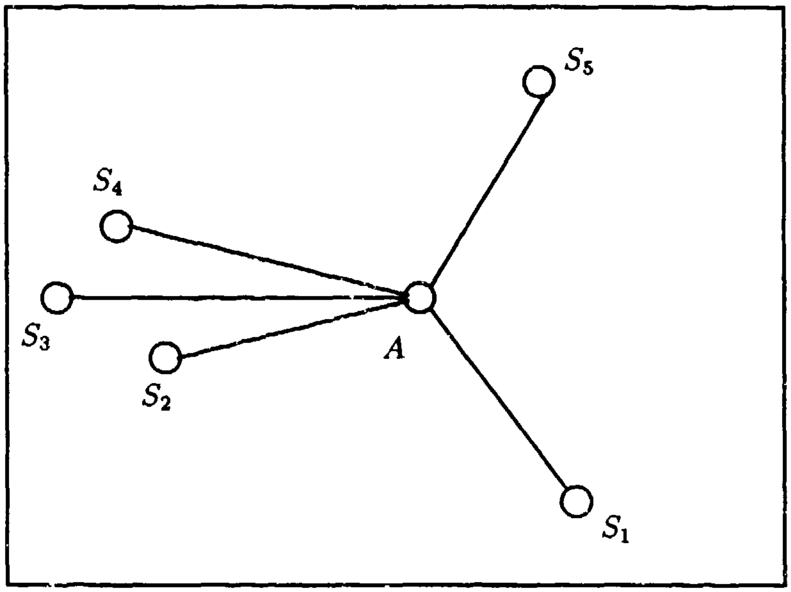

Figure 2.1. Typical Kriging Problem

distributed throughout the region and have corresponding values of $X_1 - X_5$ which are known for some specific attribute. The problem in kriging is to predict the value of $\hat{X}_0$ at the point A from the known values in the vicinity (1, 353). The value of $\hat{X}$ which minimizes the estimation error is determined by solving a set of linear equations.

Kriging Equations. In kriging, the estimate for an unknown value at a location is determined by a weighted average of sample values with the sample values in closer proximity having more weight than points further away (3, 99). Specifically, the equation for the estimator is:

$$\hat{X} = w_1 X_1 + w_2 X_2 + w_3 X_3 + \ldots + w_n X_n$$

where $\hat{X}$ is the estimator, w_1, w_2, $w_3, \ldots$, w_n are the weights, and X_1, X_2, X_3, ..., X_n are the sample values (3, 99).

If the weights sum to one and there is no trend, then the estimator is unbiased (3, 99). "This means that over a lot of estimations the average error will be zero" (3, 99). Because the estimator is a linear combination of the sample values, it is considered to be a "linear" estimator (3, 99). There are an infinite number of estimators in the above form which are unbiased and linear (3, 104). There is, however, a unique combination that will give minimum estimation error (6, 385).

To determine the combination which minimizes the estimation error, the estimation variance must be defined. The estimation variance for the general unbiased linear estimator is:

$$\sigma_\epsilon^2 = 2\sum_{i=1}^{n} w_i\bar{\gamma}(S_i, A) - \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \bar{\gamma}(S_i, S_j) - \bar{\gamma}(A, A)$$

where σ_ϵ^2 is the estimation variance, w_i and w_j are the weights, S is the sample set, A is the point to be estimated, and $\bar{\gamma}(S, A)$ is known as the average semivariogram. The semivariogram is a function describing the expected difference in value between pairs of samples with a given spatial relationship (3, 11) and will be discussed at length in the structural analysis section of this review. For any given set of observations the variance is a function only of the values of the weights. Therefore, to minimize the estimation variance, the partial derivatives of the estimation variance with respect to the weights must be set to zero and the weights must be determined by solving the resulting system of equations. To maintain the unbiased nature of the estimate, an equation must be added to the system to ensure that the weights sum to one (3, 105). Additionally, a Lagrangian Multiplier (λ) is used to ensure that the number of unknowns and the number of equations are equal. The result is the following system

of linear equations (referred to as the kriging system):

$$\begin{array}{ccccccccccccc}
w_1\bar{\gamma}(S_1,S_1) & + & w_2\bar{\gamma}(S_1,S_2) & + & \cdots & + & w_n\bar{\gamma}(S_1,S_n) & + & \lambda & = & \bar{\gamma}(S_1,A) \\
w_1\bar{\gamma}(S_2,S_1) & + & w_2\bar{\gamma}(S_2,S_2) & + & \cdots & + & w_n\bar{\gamma}(S_2,S_n) & + & \lambda & = & \bar{\gamma}(S_2,A) \\
w_1\bar{\gamma}(S_3,S_1) & + & w_2\bar{\gamma}(S_3,S_2) & + & \cdots & + & w_n\bar{\gamma}(S_3,S_n) & + & \lambda & = & \bar{\gamma}(S_3,A) \\
\ldots & + & \ldots & + & \ldots & + & \ldots & + & & = & \ldots \\
\ldots & + & \ldots & + & \ldots & + & \ldots & + & & = & \ldots \\
w_1\bar{\gamma}(S_n,S_1) & + & w_2\bar{\gamma}(S_n,S_2) & + & \ldots & + & w_n\bar{\gamma}(S_n,S_n) & + & \lambda & = & \bar{\gamma}(S_n,A) \\
w_1 & + & w_2 & + & \ldots & + & w_n & & & = & 1
\end{array}$$

In clarification of notation, the above system of equations is specified in Davis (6, 385–386) and Henley (9, 26) in the following manner:

$$\begin{array}{ccccccccccccccc}
w_1\gamma(h_{11}) & + & w_2\gamma(h_{12}) & + & w_3\gamma(h_{13}) & + & \cdots & + & w_n\gamma(h_{1n}) & + & \lambda & = & \gamma(h_{1p}) \\
w_1\gamma(h_{21}) & + & w_2\gamma(h_{22}) & + & w_3\gamma(h_{23}) & + & \cdots & + & w_n\gamma(h_{2n}) & + & \lambda & = & \gamma(h_{2p}) \\
w_1\gamma(h_{31}) & + & w_2\gamma(h_{32}) & + & w_3\gamma(h_{33}) & + & \cdots & + & w_n\gamma(h_{3n}) & + & \lambda & = & \gamma(h_{3p}) \\
\ldots & + & \ldots & + & \ldots & + & & + & \ldots & + & & = & \ldots \\
\ldots & + & \ldots & + & \ldots & + & & + & \ldots & + & & = & \ldots \\
w_1\gamma(h_{n1}) & + & w_2\gamma(h_{n2}) & + & w_3\gamma(h_{n3}) & + & \ldots & + & w_n\gamma(h_{nn}) & + & \lambda & = & \gamma(h_{np}) \\
w_1 & + & w_2 & + & w_3 & + & \ldots & + & w_n & + & 0 & = & 1
\end{array}$$

The above form of the kriging equations are equivalent to the previous equations. In the second notation, $\gamma(h_{ij})$ is the semivariogram value for the distance h_{ij} between the observations i and j. The only difference in the two systems of equations is that the first set is referring to block kriging (denoted by the $\bar{\gamma}$ structure) and the second set refers to point kriging (denoted by the γ structure) (9, 26). Block and point kriging are only two of the forms of kriging and are discussed in more detail in the section on types of kriging.

Kriging Assumptions. In point and block kriging, weak stationarity is assumed. Weak stationarity implies that all random variables X_k have the same mean, variance and autocorrelation function. This assumption is based on two conditions:

1) the expected value of the regionalized variable $Z(x)$ is the same all over the field of interest; and, 2) the spatial covariance of the regionalized variable $Z(x)$ is the same all over the field of interest (5, 92). Therefore, the expected value, $E[Z(x)] = m$ and the covariance, $E\{[Z(x) - m][Z(x + h) - m]\} = K(x, x + h) = K(h)$ where h is a vector in R_n. Strict stationarity is assumed if, in addition to the above conditions, the higher-order moments remain the same (1, 315). However, it is generally accepted in practial applications that weak stationarity is sufficient (1, 315). While this assumption is essential to point kriging, methods have been developed to accomodate violations of weak stationarity. The following section discusses a few of the methods which are used when the weak stationarity assumption is relaxed.

Types of Kriging. Henley provides a description of several kriging techniques to include point, block, lognormal, disjunctive, and universal kriging (9, 25–31). Additionally, Barnes and Johnson discuss the technique of positive kriging (2, 231–243) and Journel and Huijbregts develop the theory for cokriging (13, 324-326). "The kriging techniques are all related, and are refined versions of the weighted moving average techniques used by Krige" (9, 25). The selection of the most appropriate form of the kriging methods depends primarily on validation of the basic assumptions of kriging. Table 2.1 is adapted from a figure in *Nonparametric Geostatistics* and provides an overview of which kriging methods to use when the assumptions are not satisfied (9, 27). Several of these methods are discussed below in more detail.

Point Kriging. Point kriging was discussed previously in the development of the kriging equations. This form of kriging is used when the stationarity assumption is valid and the distribution of the ore, or measurement, is assumed normal. Davis discusses this simplest form of kriging in the context of punctual kriging and provides an example to illustrate the mechanics of the kriging system. The following example is adapted from *Statistics and Data Analysis in Geology* and demonstrates the use of kriging in estimating the water elevation at an unsampled

Table 2.1. Available Kriging Methods

DISTRIBUTION	STATIONARITY			
Normal	Simple kriging (point or block)	Universal kriging	Generalized covariances	?
Simple known (e.g. lognormal)	Lognormal kriging	?	?	?
Complex	Disjunctive kriging	?	?	?

Table 2.2. Water Table Elevation Data

Location	X Coordinate	Y Coordinate	Water Table Elevation
Well 1	3.0	4.0	120
Well 2	6.3	3.4	103
Well 3	2.0	1.3	142
Point p	3.0	3.0	

location (6, 386–390).

The basic problem is to estimate the water elevation at some point p based on the elevations at three other points in the general vicinity. The coordinates and the water table elevations at these points are listed in Table 2.2. A structural analysis was performed and determined the semivariogram for the neighborhood of $20km$ to be linear with a slope of $4.0m^2/km$.

The objective is to determine the weights, w_1, w_2, and w_3 which minimize the estimation error. After solving the kriging system of equations to determine the weights, the estimate of the water elevation $\hat{X}_p$ at the location p is obtained using

the equation $\hat{X}_p = w_1 X_1 + w_2 X_2 + w_3 X_3$. The kriging system used to determine the weights is:

$$\begin{bmatrix} \gamma(h_{11}) & \gamma(h_{12}) & \gamma(h_{13}) & 1 \\ \gamma(h_{21}) & \gamma(h_{22}) & \gamma(h_{23}) & 1 \\ \gamma(h_{31}) & \gamma(h_{32}) & \gamma(h_{33}) & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ \lambda \end{bmatrix} = \begin{bmatrix} \gamma(h_{1p}) \\ \gamma(h_{2p}) \\ \gamma(h_{3p}) \\ 1 \end{bmatrix}$$

Using the distance between the points h and the equation for the variogram, $\gamma(h_{ij}) = 4.0 * h$, the above equations are rewritten as:

$$\begin{bmatrix} 0 & 12.2 & 11.5 & 1 \\ 12.2 & 0 & 18.1 & 1 \\ 11.5 & 18.1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ \lambda \end{bmatrix} = \begin{bmatrix} 4.0 \\ 12.1 \\ 7.9 \\ 1 \end{bmatrix}$$

Solving these equations produces the following estimates for the weights:

$$\begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ \lambda \end{bmatrix} = \begin{bmatrix} 0.5954 \\ 0.0975 \\ 0.3071 \\ -0.7298 \end{bmatrix}$$

Therefore, the elevation at p is determined as $\hat{X} = 0.5954(120) + 0.0975(103) + 0.307(142) = 125.1$ meters. This example demonstrates simple or point kriging. The system of equations is appropriate for stationary data which follow a normal distribution. However, these equations must be modified when trend, or drift, is present.

Universal Kriging. Universal kriging is used when trend is present. Typically, a nonstationary regionalized variable is composed of drift, or the expected value of the variable in a neighborhood, and the residual which is the difference between the drift and the actual value (6, 393). In this form of kriging, the drift is

removed from the regionalized variable and the stationary residuals are kriged. In short,

> Universal kriging can thus be regarded as consisting of three operations: First, the drift must be estimated and removed. Then, the stationary residuals are kriged to obtain needed estimates. Finally, the estimated residuals are combined with the drift to obtain estimates of the actual surface (6, 393).

The drift is generally represented by a first or second-order polynomial. One method for removing the drift is to estimate the polynomial for the drift during the structural analysis and then subtracting this drift from the data. However, the complexities of combining the neighborhood size with the drift equations generally prohibits this method in practice. Fortunately, the equations for the drift can be incorporated directly in the kriging equations.

Davis (6, 394–395) provides the matrix form of the universal kriging system when the first-order-polynomial drift at a point p is defined as:

$$M_P = \alpha_1 X_{1i} + \alpha_2 X_{2i}$$

In this equation, the α's are drift coefficients which must be estimated and X_{1i} and X_{2i} are the coordinates of the ith control point.

For consistency, the matrix form for universal kriging is presented for the problem demontrated above in *Point Kriging.* The equations are as follows:

$$\begin{bmatrix} \gamma(h_{11}) & \gamma(h_{12}) & \gamma(h_{13}) & 1 & X_{11} & X_{21} \\ \gamma(h_{21}) & \gamma(h_{22}) & \gamma(h_{23}) & 1 & X_{12} & X_{22} \\ \gamma(h_{31}) & \gamma(h_{32}) & \gamma(h_{33}) & 1 & X_{13} & X_{23} \\ 1 & 1 & 1 & 0 & 0 & 0 \\ X_{11} & X_{12} & X_{13} & 0 & 0 & 0 \\ X_{21} & X_{22} & X_{23} & 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ \lambda \\ \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} \gamma(h_{1p}) \\ \gamma(h_{2p}) \\ \gamma(h_{3p}) \\ 1 \\ X_{1p} \\ X_{2p} \end{bmatrix}$$

While universal kriging solves the problem of nonstationarity, other forms provide the means for violations of normality.

Lognormal Kriging. The distribution of ore grades, or the regionalized variable, often is not normally distributed. In some cases, a better representation of the data is found by fitting a lognormal distribution. In lognormal kriging, the variable is transformed by the equation $y_i = log(x_i + a)$ where a is an arbitrary constant used to optimize the fit to a normal distribution (9, 28). Using this transformation, the semivariograms and kriging estimates are obtained. However, the resulting estimates are in terms of logarithms and must be converted using an inverse transformation. Unfortunately, the inverse transformation does not produce a linear estimate of the value being estimated at a point p and the estimation variance is not necessarily minimized (9, 28)

Disjunctive Kriging. Although disjunctive kriging is beyond the scope of this study, the technique is briefly described to illustrate the potential of kriging when the regionalized variable is not normally distributed. In simple kriging, the estimate of the value at a point p is determined by a linear combination of the values at other points in the neighborhood. Theoretically, the best estimator is the conditional expectation of the value at p based on the neighboring values and is linear when the distribution is normal and stationary (9, 29). Without the assumptions of normality and stationarity, the best estimator may not be linear. In disjunctive kriging, the data is transformed to a normal distribution. Journel and Huijbregts describe this transformation using a set of Hermite Polynomials (13, 573-580). The result of this technique is a nonlinear estimator for the value at the point p.

Structural Analysis

The semivariogram was briefly mentioned earlier in this review. Yakowitz and Szidarovszky note:

> The kriging method is composed of two activities, (i) inferring the variogram from the data and (ii) assuming that the inferred variogram is indeed exact, providing a best linear unbiased estimator and associated error variance (22, 23–24).

Thus far, this review has concentrated on the second activity. To complete the review of kriging, the first activity will be discussed.

Definition. Journel and Huijbregts emphasize that the first and most important step in any geostatistical study is structural analysis (13, 12). "Structural analysis is the name given to the procedure of characterizing the structures of the spatial distribution of the variables considered (e.g., grades, thicknesses, accumulations)" (13, 12). Journel and Huijbregts further explain how the variogram quantifies and summarizes the structural information and then serves to interject this knowledge into the geological study (13, 12).

The Variogram. In geostatistics, the three second-order moments considered are the variance, the covariance, and the variogram (13, 31). According to Omre, "the variogram function is the backbone of geostatistical analysis" (17, 107). Basically, the variogram function is defined as the variance of the increment $[Z(x_1) - Z(x_2)]$. This increment is written as:

$$2\gamma(x_1, x_2) = Var\{Z(x_1) - Z(x_2)\}.$$

The semivariogram is simply $\gamma(x_1, x_2)$.

In practice, only an *estimator* of the theoretical variogram is available. This estimator is known as the experimental variogram and is calculated as follows:

$$2\gamma^*(h) = \frac{1}{N'}\sum_{i=1}^{N'}[z(x_i + h) - z(x_i)]^2$$

where N' is the number of pairs of data values at a distance of h apart from one another.

The next step in the structural analysis procedure is to fit a model to the experimental variogram. David proposes several methodologies for performing this task (5, 119–120). While this procedure is sometimes considered an art, one approach suggests selecting a model and then determining the parameters through numerical

least-squares fitting (5, 119). Cressie proposes minimizing a weighted sum of squares and indicates that work by Zimmerman and Zimmerman shows that the weighted least squares approach never performs poorly and usually does well (4, 198). The process of fitting models to experimental variograms is not trivial. Therefore, this review will only discuss a few of the various models which appear in the literature.

Standard Models. Three of the more common models are the linear model, the De Wijsian model, and the spherical model (5, 120–122). A brief introduction to each of these models is provided.

Linear Model. The equation for the linear model is of the form $\gamma(h) = ah + b$. This is one of two models used in practice which does not have a sill (5, 120). The sill is defined as the variance of the samples and will be mentioned again in the discussion of the spherical model. David suggests that a visual fit of the data is usually satisfactory and that a least-squares method could be used. However, he emphasizes that weighted least-squares should be used since the number of pairs used in calculating the variogram decreases as h increases (5, 120).

De Wijsian Model. The form for the De Wijsian model is $\gamma(h) = a\ln(h)+b$. However, "one usually writes $a = 3\alpha$ and calls it the coefficient of intrinsic dispersion" (5, 121). This model is the second of two commonly used models which does not have a sill. This model is named after Prof. H.J. de Wijs and is used when the experimental data plots as a straight line on a logarithmic scale (5, 120). Weighted least-squares is also appropriate for fitting this model.

Spherical Model. (5, 80) The spherical model is the most common model and is defined by three parameters: a, C, and C_0. The first parameter, a, is called the range and is used to determine the zone of influence. The sampling error is known as the nugget effect and is denoted by C_0. Finally, the third parameter, C, is used in conjunction with C_0 to determine the sill, $(C + C_0)$. The form of the

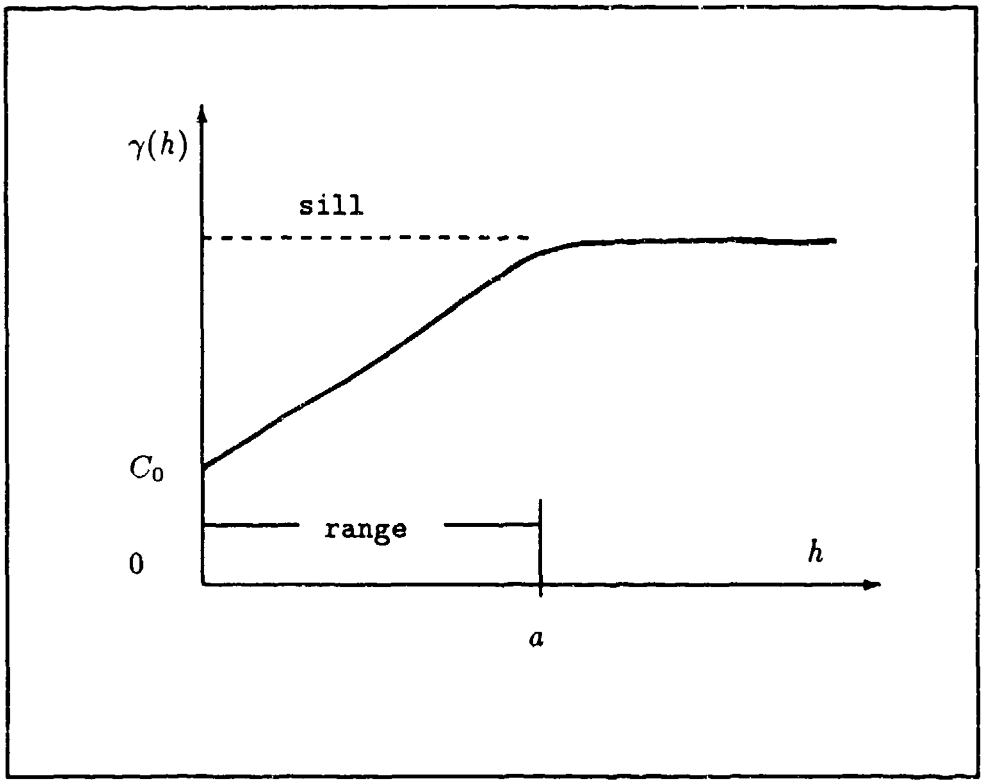

Figure 2.2. Spherical Model Variogram

spherical model is as follows:

$$\gamma(h) = \begin{cases} C(\frac{3}{2}\frac{h}{a} - \frac{1}{2}\frac{h^3}{a^3}) + Co & \text{if } h < a \\ C + C_0 & \text{if } h \geq a \\ 0 & \text{if } h = 0 \end{cases}$$

The shape of this model is shown in Figure 2.2 adapted from *Geostatistical Ore Reserve Estimation* (5, 102).

David provides a detailed example for fitting the spherical model (5, 122–125). In addition to a weighted least-squares approach, David suggests the following:

> One usually starts by fitting the tangent at the origin to the curve, the intercept of that tangent at the origin is the nugget effect C_0. The sill of the variogram is equal to the variance of the samples in the deposit,

which can be computed from the samples. This defines $C + C_0$; finally the tangent at the origin intersects the sill at a distance which is equal to $\frac{2}{3}a$. This defines a (5, 122).

Problems with Anisotropy. Anisotropies are typically classified in one of two categories: geometric (or affine) and zonal (or stratified) (5, 134). Geometric anisotropy refers to the situation where the value or expected variation varies more quickly in one direction than in another. An example of geometric anisotropy that occurs in geology is the situation where the unit of measure varies over the region of interest. This type of anisotropy can be handled by adjusting the coordinates of the data sets or by using different variograms for different directions. Zonal isotropy is characterized by qualitative variations or separations of the data into zones. This form is difficult to treat. An example of zonal isotropy in geology is the situation where different rocks are sharply divided within sediment beds (9, 98–99).

The anisotropy ratio (or affinity modulus), k, is equivalent to the change in distance units between axes. For example, if the same relationship exists between points 50 feet apart in one direction and 300 feet apart in another direction, then k is equal to 300/50 (5, 134–135). David suggests that geometric isotropies are easily recognized after plotting the variograms in two directions. For the linear, De Wijsian, and spherical models, the variation for a distance h in one direction is equivalent to the variations for a distance kh in another. Therefore, the equations for the two variograms for direction 1 and direction 2 for these three models are presented as follows: for the linear model,

$$\gamma_1(h) = ah$$

$$\gamma_2(h) = akh$$

for the De Wijsian model,

$$\gamma_1(h) = a\ln(h) + b$$

$$\gamma_2(h) = a\ln(kh) + b$$

and, for the spherical model,

$$\gamma_1(h) = C[\frac{3h}{2a} - \frac{h^3}{2a^3}] + C_0$$

$$\gamma_2(h) = C[\frac{3kh}{2a} - \frac{k^3h^3}{2a^3}] + C_0$$

Proceeding further, the distance h can be decomposed into two components h_1 and h_2 corresponding to directional axis. Therefore, the distance between the two points (x_1, y_1) and $(x_2.y_2)$ is $h' = \sqrt{(x_1 - x_2)^2 + k^2(y_1 - y_2)^2}$. This distance measure is used in obtaining the variogram equation for more than one direction (5, 138). In this presentation, two perpendicular directions are considered. Treating geometric anisotropy when the axes are not perpendicular and the treatment of zonal anisotropy is beyond the scope of this study. For more information reference David (5, 134–148).

Bayesian Statistics

Bayesian statistics is the branch of statistics characterized by the use of prior information. This section concentrates on only one area in the broader class of Bayesion statistics, the Kalman filter. The definition, assumptions, and equations are provided.

Definition of a Kalman Filter. "A Kalman filter is simply an optimal recursive data processing algorithm" (15, 4). Maybeck points out that the Kalman filter is "optimal with respect to virtually any criterion that makes sense" and that one aspect of this optimality is that the Kalman filter uses all available information. The filter is recursive from the standpoint that it "does not require all previous data to be kept in storage and reprocessed every time a new measurement is taken" (15, 4). The model is classified under Bayesian Statistics because the filter propagates the probability density of the values conditioned on the knowledge of the data being measured.

Assumptions. The formulation of a Kalman filter includes the validation of three basic assumptions. The first assumption is that the system model is linear (15, 7). Secondly, the measurement noises are not correlated in time. Maybeck refers to this quality as "whiteness" and explains that this means "if you know what the value of noise is now, this knowledge does you no good in predicting what its value will be at any other time" (15, 7). (The analogy of space and time within the realm of spatial statistics is apparent. Further analogy is drawn with the kriging assumptions by David in the previous section.) The third assumption of a Kalman filter is that these measurement noises are Gaussian. The basis for this assumption is that measurement noise includes the effects of many sources and, when added together, resemble the Gaussian probability density.

Equations. This review is concerned primarily with the application of a relatively basic form of the Kalman filter model. The following equations are adapted from the example provided by Maybeck in the introduction of *Stochastic Models, Estimation, and Control. Vol 1.* (15, 9–15).

Let z_1 be a measure at time t_1 and $\sigma_{z_1}^2$ the variance of this measure. In Maybeck's example, this measure refers to an estimate of the location at a particular time. However, this measure could also be the first measurement of the value of a surface point in a facial region. If the actual value at t_1 is $x(t_1)$, then the conditional probability of x_{t_1} conditioned on z_1 is $f_{x(t_1)|z(t_1)}(x|z_1)$. Therefore, the best estimate of $x(t_1)$ is $\hat{x}(t_1) = z_1$ and the error variance is $\sigma_x^2(t_1) = \sigma_{z_1}^2$.

Furthermore, let z_2 be the measure of $x(t_2)$ at time t_2 and $\sigma_{z_2}^2$ be the variance of z_2. It can be shown that the conditional density of $x(t_2)$, given z_1 and z_2, is a Gaussian density with mean μ and variance σ^2 where

$$\mu = [\sigma_{z_2}^2/(\sigma_{z_1}^2 + \sigma_{z_2}^2)]z_1 + [\sigma_{z_1}^2/(\sigma_{z_1}^2 + \sigma_{z_2}^2)]z_2$$

and,

$$1/\sigma^2 = (1/\sigma_{z_1}^2) + (1/\sigma_{z_2}^2)$$

The best estimate of $x(t_2)$ is $\hat{x}(t_2) = \mu$ with variance σ^2. Realizing that $\hat{x}(t_2)$ includes all the information in $f_{x(t_2)|z(t_1),z(t_2)}(x|z_1, z_2)$, further estimates can be obtained at time t_n based on known information at t_{n-1}. Therefore, the following equations provide the best estimate of a value $x(t_n)$ and the variance $\sigma_x^2(t_n)$:

$$\hat{x}(t_n) = \hat{x}(t_{n-1}) + K(t_n)[z_n - \hat{x}(t_{n-1})]$$

and

$$\sigma_x^2(t_n) = \sigma_x^2(t_{n-1}) - K(t_n)\sigma_x^2(t_{n-1})$$

where

$$K(t_n) = \sigma_{z_{n-1}}^2 / (\sigma_{z_{n-1}}^2 + \sigma_{z_n}^2)$$

Multivariate Analysis

In general, multivariate analysis is the application of methods which deal with the simultaneous relationships of several variables charactersitic of objects in a sample. The use of multivariate analysis in a variety of fields is firmly established and well documented. Davis presents several uses for multivariate methods in geology (6, 468–615). Futhermore, Flury and Riedwyl demonstrate the use of a multivariate technique, principle components analysis, in the analysis of anthropometric data to support the sizing of protective masks for the Swiss army (8, 218–228). Two of the more commonly used methods in multivariate analysis are factor analysis and cluster analysis. These methods are introduced below.

Factor Analysis. Basically, factor analysis is a method for reducing the number of variables in a data set by determining a new and smaller set of variables which accounts for the common variation. Typically, the new set of variables consists of factors which represent the true dimensionality of the data. These factors are determined by analyzing the interrelationships among the variables. Dillon and Goldstein define factor analysis as follows:

Factor analysis attempts to simplify complex and diverse relationships that exist among a set of observed variables by uncovering common dimensions or factors that link together the seemingly unrelated variables, and consequently provides insight into the underlying structure of the data (7, 53).

Cluster Analysis. Cluster analysis refers to a number of techniques which classify objects in homogeneous and distinct groups. The definition of a cluster is often determined by the researcher. The goal ,however, is to determine groups of observations which have some defined similarity. Dillon and Goldstein discuss several of the partitioning and hierarchical techniques and present a listing of the more commonly used clustering programs (7, 167–207).

Summary

Several kriging and structural analysis topics from the literature were presented. Specifically, this paper discussed the origin of kriging, defined kriging in terms of being a best linear unbiased estimator, and presented the kriging system of equations. Additionally, various forms of kriging and several of the more commonly used models for the theoretical variogram were discussed. This review presented the background for kriging as documented in the field of geostatistics. Other topics presented in this review include the introduction of Kalman filtering, factor analysis, and cluster analysis. These topics were discussed to the degree that they were used in this study.

III. Methodology

This chapter presents the methodology used in completing the objectives outlined in Chapter I. The first section discusses the procedures used in collecting and orientating the data. The second section develops the kriging procedure for estimating facial surfaces. This section is followed by the development of a recursive procedure for aggregating the individual estimates obtained by the kriging methodology. The fourth section documents the application of the kriging and updating procedures to the surface area which will influence the design of the night- vision goggles. Finally, the last section discusses the multivariate analysis procedures investigated for use in clustering the faces.

Data Collection and Orientation

Because the majority of the data manipulation was completed prior to the start of this study, data collection and orientation were not listed as objectives of this study. However, the processes of collecting and preprocessing the data were critical steps in the analysis performed in this thesis. Therefore, the procedures for data collection and data orientation are included in the methodology.

Data Collection. As discussed in the introduction, researchers at AAMRL have the ability to collect 3-dimensional facial data with the use of a laser scanner. The manner in which the data is recorded is complex and, to some degree, beyond the scope of this discussion. However, a brief explanation is appropriate. The laser scanner is mounted on a rotating arm which rotates around the head of the subject while the subject is seated calmly in a chair. The measurements of 131072 points are recorded for each subject as the scanner rotates. The number of points is determined by the 512 locations on the x axis and 256 locations on the y axis. The x axis refers to axis of rotation (angle) and the y axis refers to the altitude of the point from

Table 3.1. Anthropometric Landmark Names

Landmark	Name	Landmark	Name
1	Right Frontotemporale	10	Right Infra Malar
2	Glabella	11	Pronasale
3	Right Zygofrontale	12	Subnasale
4	Right Tragion	13	Right Chelion
5	Right Zygion	14	Stomion
6	Right Ectocanthus	15	Right Gonion
7	Sellion	16	Promenton
8	Right Infraorbitale	17	R. Midlateral Infra Mandibular
9	Right Infra Zygion	18	Menton

the x axis. The third coordinate of the points, the z values, are determined by measuring the depth of the point at each x, y location. The depth is calculated using a triangulation procedure based on the reference point on the scanner, the point being measured, and an arbitrary (but constant) point in the environment surrounding the subject. In essence, the relative position of the points to each other is being measured. Prior to each scanning, a mark is placed on selected landmarks. Several of these landmarks are identified in Figure 3.1 and defined in Table 3.1. The marks do not reflect the light source and are omitted from the initial data base. However, these points are reinserted after the scanning process during a graphical review of the data. At the same time, the landmarks which were not marked are identified and recorded. As a final step in the process, the data is transformed to an axis system based on the landmark coordinates (20).

Data Orientation. After the data was collected for each of the subjects, the sets had to be aligned on a common coordinate system. There were two reasons which necessitated this orientation of the data sets. First, the individuals could not possibly all sit in the same position with their heads in the same orientation while being scanned. Secondly, the goal of the study was to estimate the facial surfaces

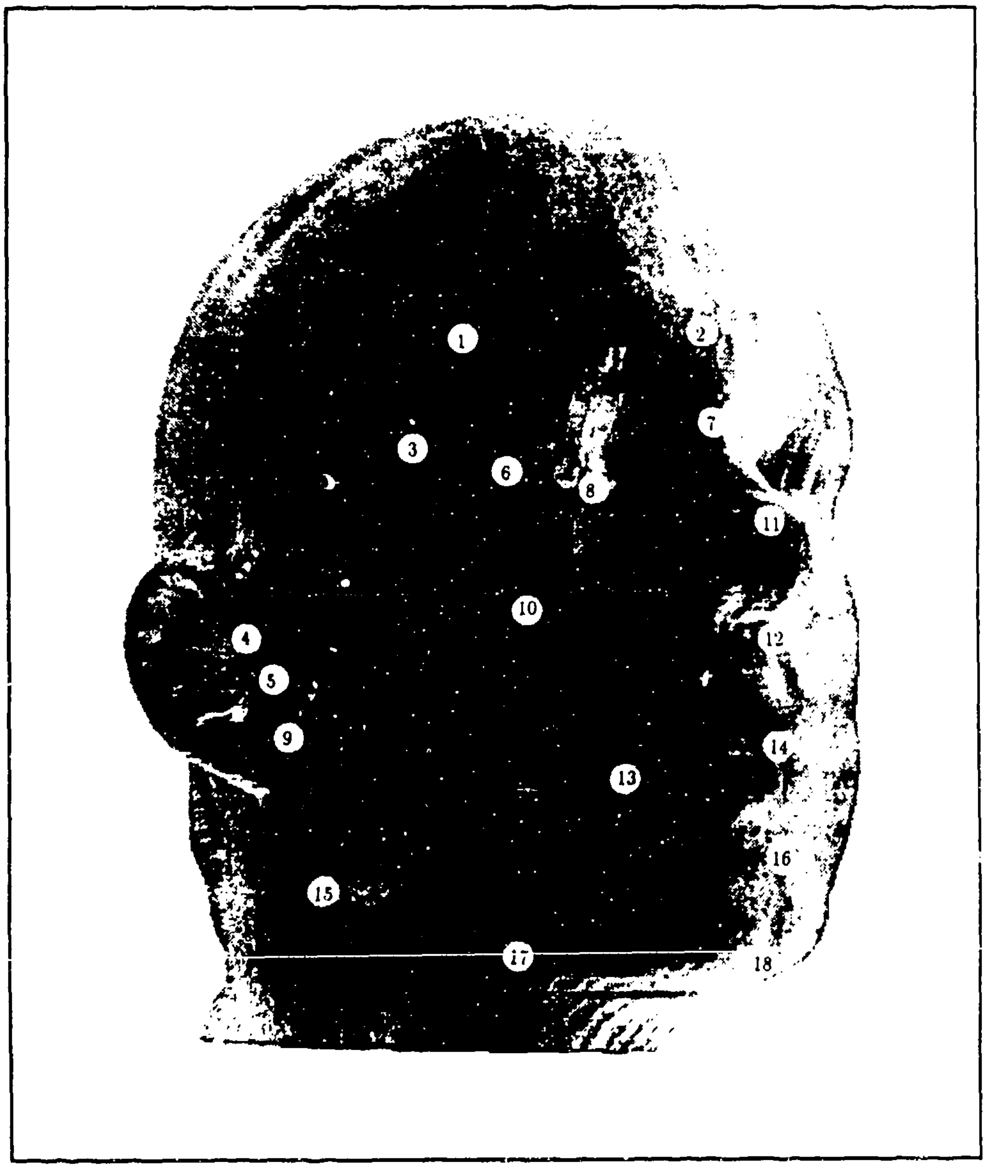

Figure 3.1. Anthropometric Landmarks

which minimize the variability between subjects. To achieve this goal, the data for each subject had to be aligned in a manner which minimized the differences in the reference points for the landmarks. In other words, the attempt was to align the faces so that all the features were in the same relative position and orientation.

The alignment was achieved using a multivariate, nonlinear optimization routine implemented through a program written by Dr. David G. Robinson. This program identifies four landmarks for each data set and minimizes the distance from these points to four standard points established at positions which were chosen to force the correct orientation of the faces. These four landmarks were also used in truncating the data sets to include only the points within the region bounded by these points. Additionally, the x axis was transformed to rectangular coordinates so that the data was represented in a rectangular grid structure as opposed to the spherical structure inherent in the collection process.

Several programs were used to produce the configuration of the data used in this thesis. These programs were developed by Dr. Robinson, Joyce Robinson (Scientific Research Laboratories), and Dr. Robert Beecher(Beecher Associates).

Procedure Development

As discussed previously in the literature review, kriging involves both the estimation of the variogram through the structural analysis of the data and the determination of the estimates and error variances. In developing a kriging procedure for the anthropometric data, these two activities were treated as separate tasks. The first task is discussed below under Structural Analysis and the second task is considered under the heading of *Kriging.*

Structural Analysis. Structural analysis is key to the implementation of kriging in any field. This analysis must validate the assumptions of stationarity and isotropy or correct for any shortfalls inherent to the data. For this study, the struc-

tural analysis consisted of removing the global trend from the data, calculating the experimental variograms for each subject, and estimating the parameters of three commonly used variogram models.

Trend Analysis. Some nonstationarity is inherently present in facial data. The removal of this trend was accomplished in two steps: 1) the removal of the global trend by differencing each subject data set with the average of 30 data sets, and 2) the inclusion of first-order terms in the kriging equations to account for local stationarity within the region of influence. The first step is discussed in this section and the second step is presented in the kriging methodology.

To facilitate the removal of the global trend, a method for representing the data in a standardized manner was developed. The orignal data files were formatted so that each record included the x, y, and z coordinates of a data point. The number of data points varied from subject to subject, and these points were irregularly distributed in the xy plane. (Reference Figure 3.2.)

To simplify the data configuration, a grid was imposed upon the xy axes. (Reference Figure 3.3.) Within each block, the average z value was calculated and used to represent the surface value for the midpoint of the block. Using this approach, the number of data points was reduced from approximately 10000 to exactly 5000 (determined by the grid dimensions). The number of points in the data sets used in this study ranged from 7477–14861. Furthermore, the points in the new configuration were now regularly spaced at standardized grid points. This transformation of data was accomplished using the GRID program in Appendix G which determined the i,j indices corresponding to the row and column of the grid, the x,y coordinates for the midpoint of the block, the z value representing the average value of the block, the variance, and the number of points used to calculate the average of the block. The number of intervals on the x axis was 100 with values ranging from 0.0 to 4.0, while the number of intervals on the y axis was 50 with values ranging from 100.0

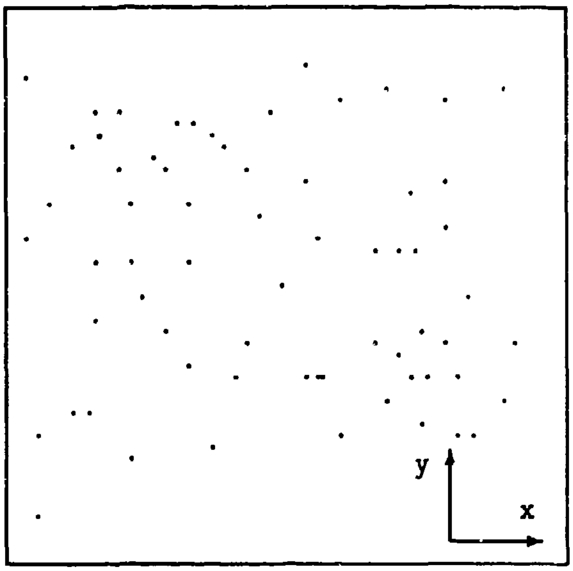

Figure 3.2. Irregularly Distributed Points

to 300.0.

Graphical representations of the transformed data were created using the Interactive Data Language(IDL) software package. The transformed data was input to another program which created an input file for an IDL procedure. Reference Appendix G. The plot in Figure 3.4 was produced using this procedure. Because of the relatively large size of the data files representing the surfaces, graphical analysis was essential in the analysis of the data. Appendix A includes the plots of all subjects used in this study.

Using the grid configuration provided a means for comparing and averaging the surface values for the subjects at common points across the region of interest. The average value at each midpoint was calculated using the data from 30 subjects. This "average" surface represents the global trend and is presented in Figure 3.5.

To remove this trend from the data, the average value for each midpoint was subtracted from the value of the midpoint for each data set. This differencing pro-

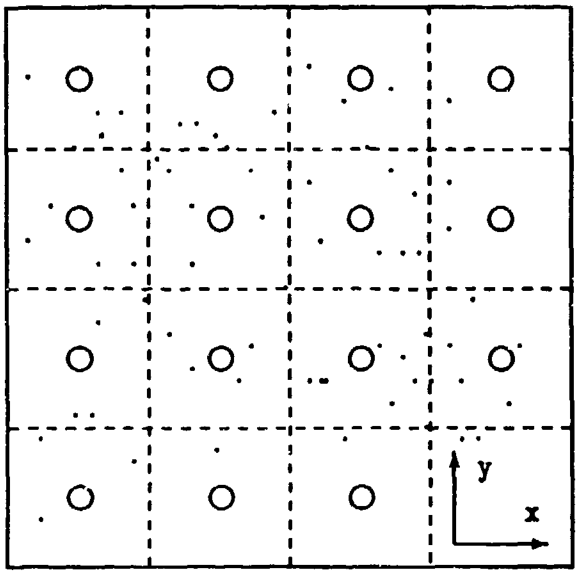

Figure 3.3. Rectangular Grid Configuration

duced a new data set consisting of residual values. A program (reference Appendix G) was used to create the residual files. Figure 3.6 is the plot of the residuals obtained from subtracting the global trend from Subject 09. Appendix B contains the residual plots for the subjects used in this study.

Experimental Variogram. The residual data sets were used to calculate the experimental variograms for each individual. The grid structure of these files simplified the variogram calculations. Typically, data are configured either at regularly spaced grid points or at irregularly distributed points throughout the region of study. The original data files configured the data in an irregularly distributed manner which would have increased the difficulty of the variogram analysis. The calculation of the variogram is simpler when data are aligned in a grid structure because the directions and distances are standardized. Specifically, the coordinates of each grid point are a multiple of some incremental distance δh; and, the standard directions, E-W, N-S, NW-SE, and SW-NE, are easily determined. In the irregu-

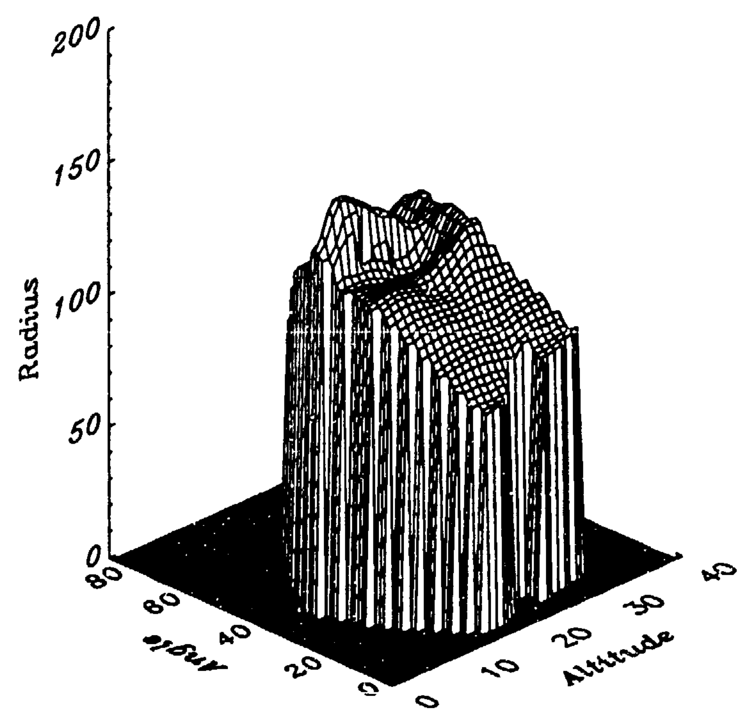

Figure 3.4. Subject 09

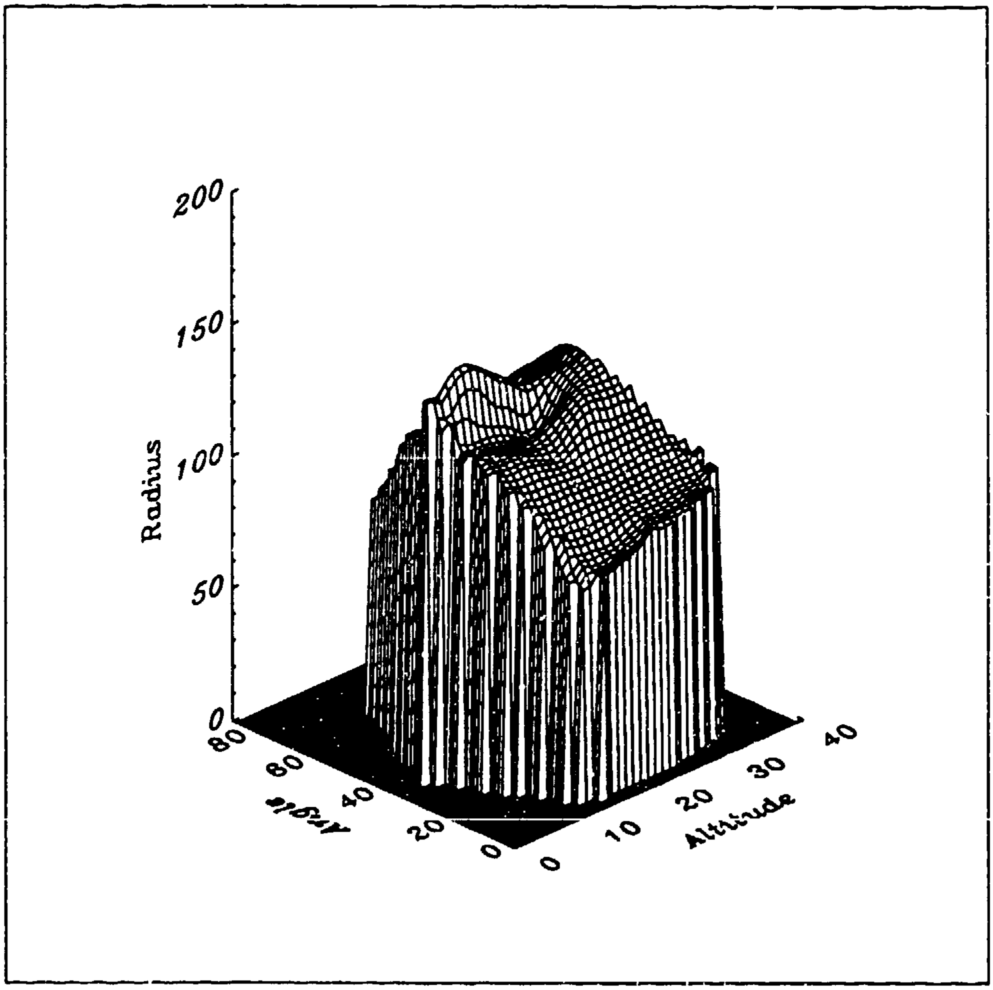

Figure 3.5. Global Trend

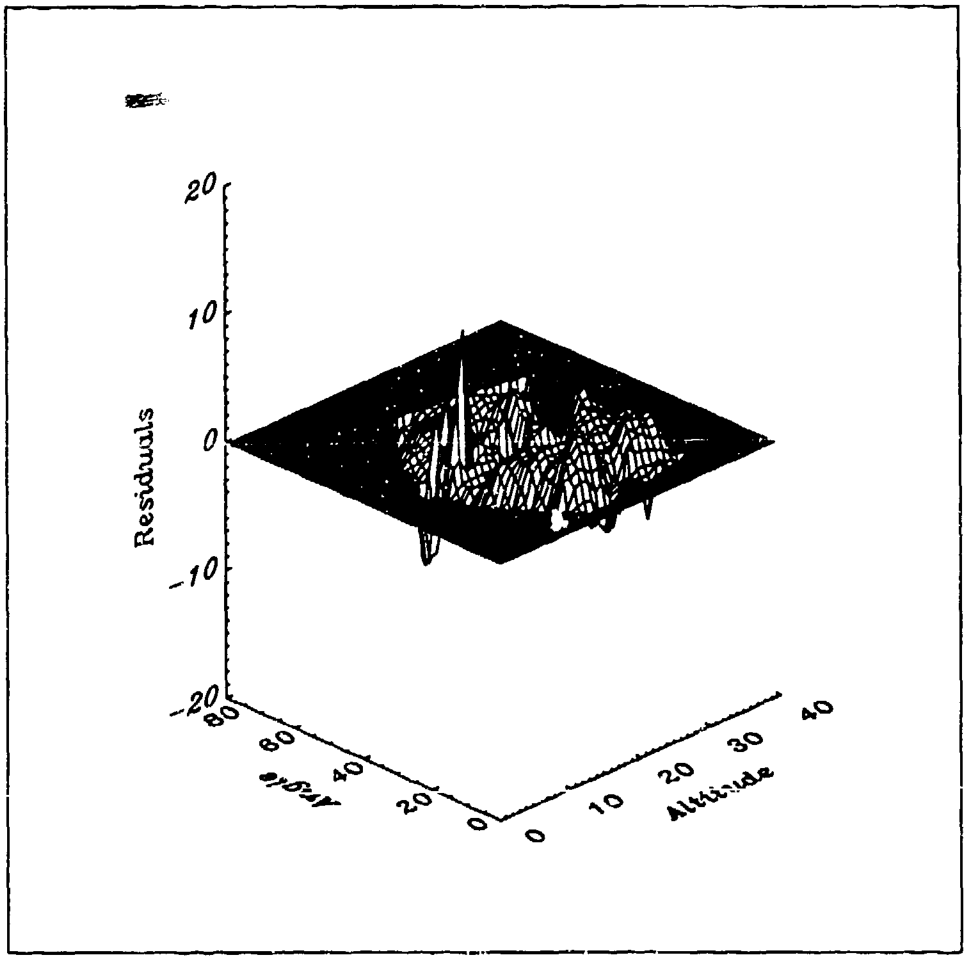

Figure 3.6. Subject 09 Residuals

larly spaced pattern, angles and distances are not unique and must be calculated independently for each set of points. Additionally, approximate directions must be determined by setting a range for the angular values.

The experimental variograms were estimated using the FORTRAN program in Appendix G. In short, this program iteratively determines the number of pairs of data points at a distance h, N', for each increment δh and for each direction. The estimator is:

$$2\gamma^*(h) = \frac{1}{N'}\sum_{i=1}^{N'}[z(x_i + h) - z(x_i)]^2$$

The variograms for each of the four directions (N-S, E-W, NW-SE, and SW-NE), for all subjects, were calculated. Figure 3.7 illustrates the four variograms for Subject 160. The $\gamma(h)$ value represents the value of the variogram at a distance h. In other words, this value represents the relationship of all the points which are at a distance h from each other. Although the function appears continuous over the range of h, (reference Figure 3.7), the experimental variogram is actually a discrete function based on the incremental units, or lags, for which the points here compared. The theoretical variogram, however, is a continuous function and is discussed later. Figure 3.7 also suggests that the data is isotropic. The variograms appear to follow the same functional pattern in all directions. This isotropic quality was not inherent to the data and was achieved by carefully defining the dimensions of the grid structure.

Several factors were considered in establishing the grid dimensions. In the trend analysis, the dimensions of the grid were restricted in that the spacing between points had to be relatively small. The analogy of pixels in a newspaper picture provides the reasoning for this restriction. The points must be close enough to characterize the shape of the the region. With respect to the variogram calculations, the dimensions were required to support a range of distance increments within the range, or zone of influence. Additionally, the relationship of δx to δy had to account for the anisotropy ratio k. Figure 3.7 represents the form of the variograms in their final form. In earlier

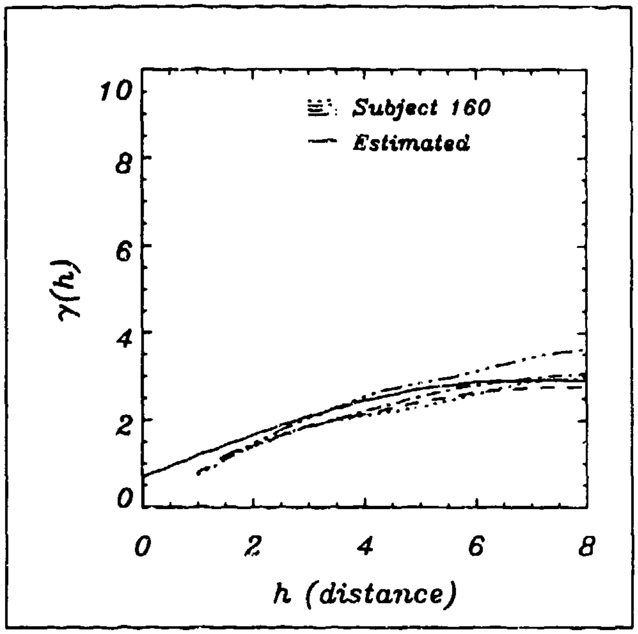

Figure 3.7. Variograms for Subject 160

attempts, ranges of h reached values of 100 to 200. Unfortunately, the variograms were dissimilar beyond distances of approximatley 10 units. Grid dimensions of 200 by 200 were also considered. However, using these dimensions, the data appeared to be geometrically anisotropic. This shortfall was corrected by changing the ratio between the x and y grid dimensions. The grid dimensions for the trend analysis and the variogram analysis were not required to be the same. However, the data analysis was simplified using the same dimensions. In the kriging analysis, a grid structure was also used. A discussion of appropriate dimensions for the kriging blocks is presented in the the kriging section.

Theoretical Variogram. A theoretical variogram was modeled using the experimental variograms calculated in the previous section. The data of the variograms was consolidated into a single data file and input to a weighted least-squares program (reference Appendix G) to obtain the parameter estimates. The following discussion presents the steps accomplished in this procedure.

Four variograms were determined for each subject— one for each direction. The goal of this analysis was to determine a single variogram which represented the relationship of the points at a distance h for the four variograms of all the subjects together. To achieve this goal, the isotropic assumption had to be verified. In general, all of the variograms would have to follow the same pattern. The sample of 30 subjects used in the trend analysis was also used in this analysis. The hypothesis that a sample of 30 would represent the true relationship for the population was assumed. Perhaps the simplest method for verifying the isotropic assumpion was to plot the variograms on the same scale. Figure 3.8 displays a plot of the variograms for four directions for the sample of 30 subjects. Based on this plot, it was determined that the variograms did not follow the same pattern and, therefore, were anisotropic. However, further analysis revealed that five subjects appeared as outliers. These five subjects were removed from the variogram analysis. Additionally, this discovery prompted the multivariate analysis to determine if the differences in the variogram structures occurred as the result of natural groupings, or sizes, of the individuals.

After removing Subjects 01, 07, 12, 89, and 150, the variograms were replotted. Reference Figure 3.9. This figure suggests that the isotropic assumption is valid in the range 0–10 for all the subjects and for all four directions. This isotropic behavior was crucial to the development of the procedures in this study. The importance of the grid dimensions must be emphasized. The variograms appeared to follow a similar pattern because the dimensions of the grid compensated for the anisotropy ratio k. In the kriging analysis, the distances were not equivalent to the block units and therefore this ratio was critical to the variogram calculations. The appropriate ratio was achieved by dividing the range of the variograms in the N-S direction by the range of the variograms in the E-W direction. To simplify the variogram calulations, the dimensions were scaled to reflect this ratio. In the kriging analysis, the ratio was incorporated into the distance calculations within the kriging programs.

The discrete data points for the experimental variogram consisted of three

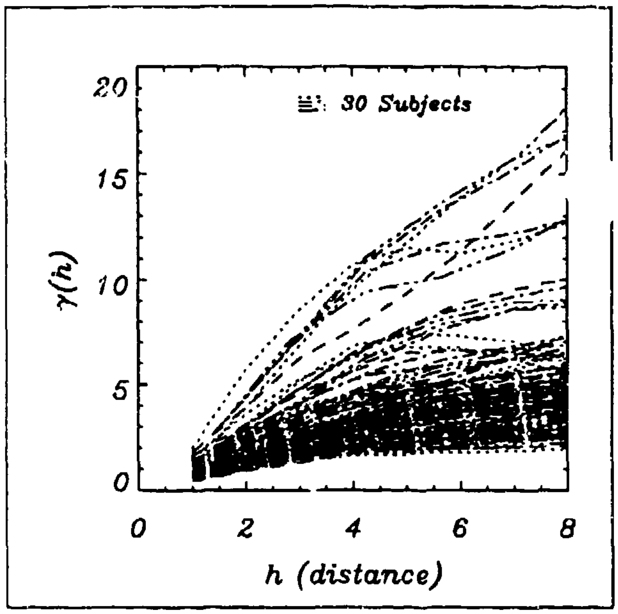

Figure 3.8. Variograms for 30 Subjects

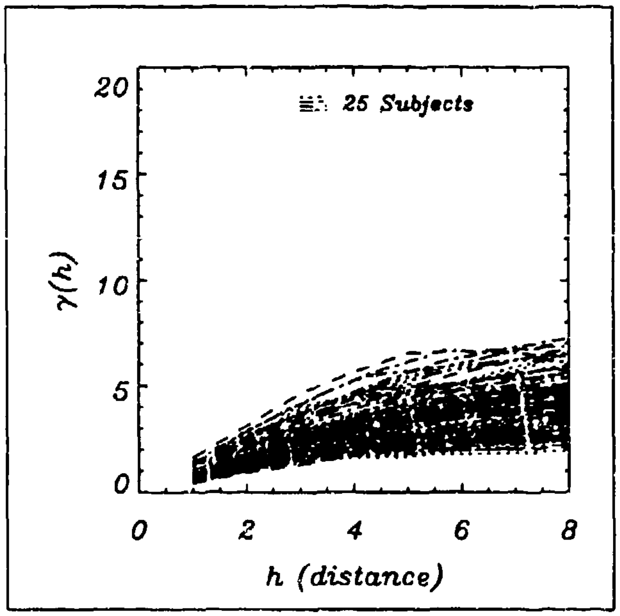

Figure 3.9. Variograms for 25 Subjects

variables: h, $\gamma(h)$, and N', the number of pairs of points used to determine $\gamma(h)$. A data file containing these variables for the variograms of the 25 subjects which matched was created using the program in Appendix G. With this data file and a regression program, the parameters of the theoretical variogram were estimated.

The regression program is provided in Appendix G. This program fits the data to the linear, De Wijsian, and spherical models presented in Chapter II. While a complete review of least-squares regression is beyond the scope of this effort, a brief description of the technique and the program is appropriate.

The weighted least-squares estimates of the regression coefficients in matrix notation are:

$$b = (X^T W X)^{-1} X^T W Y$$

where W represents the weights (16, 327). Weighted least-squares was used to account for the various number of pairs use in determining $\gamma(h)$. For this analysis,

$$W = \begin{bmatrix} \hat{w}_1 & 0 & 0 & \dots & 0 \\ 0 & \hat{w}_2 & 0 & \dots & 0 \\ \vdots & \vdots & 0 & & \vdots \\ 0 & 0 & 0 & \dots & \hat{w}_n \end{bmatrix}$$

where $\hat{w}_i$ is the number of pairs of points.

Furthermore,

$$Y = \begin{bmatrix} \gamma(h)_1 \\ \gamma(h)_2 \\ \vdots \\ \gamma(h)_n \end{bmatrix}$$

The X matrix specification was related to the model structure. For the linear

model,

$$X = \begin{bmatrix} 1 & h_1 \\ 1 & h_1 \\ 1 & \vdots \\ 1 & h_n \end{bmatrix}$$

For the De Wijsian model,

$$X = \begin{bmatrix} 1 & \ln h_1 \\ 1 & \ln h_1 \\ 1 & \vdots \\ 1 & \ln h_n \end{bmatrix}$$

Finally, for the spherical model,

$$X = \begin{bmatrix} 1 & h_1 & h_1^3 \\ 1 & h_1 & h_2^3 \\ 1 & \vdots & \vdots \\ 1 & h_n & h_n^3 \end{bmatrix}$$

To determine the regression coefficients, the normal equations were rewritten in the following form:

$$(X^T W X)b = X^T W Y$$

In this $Ax = B$ structure, the system of equations was solved using the LUDCMP and LUBKSB routines adapted from *Numerical Recipes.* The LUDCMP routine decomposes A into the product of two matrices, L and U, where L is lower triangular and U is upper triangular. LUBKSB performs backsubstitution. The combination of these programs provided an efficient method for solving the system of linear equations. (18, 29–38).

The parameters estimates for a and b for the linear model ($\gamma(h) = ah + b$) and De Wijsian ($\gamma(h) = a\ln(h) + b$) model correspond to b_1 and b_2 directly. However,

the parameters of the spherical model (C, C_0, and a) were found by the following relationships:

$$C_0 = b_1$$
$$C = \sigma^2 - C_0$$
$$a = \frac{3}{2}\frac{C}{b_2}$$

where σ^2 was the variance of the sample points.

Kriging. Having developed an estimate of the theoretical variogram, the next step in the procedural development was to estimate the points on the surface of interest and the variance associated with these estimates. Universal kriging was used for estimating these values.

As discussed in the structural analysis, the configuration of the data was an important consideration in developing the analysis techniques. The rectangular grid structure was used in the kriging analysis to determine the range and density of the surface estimate. The range was chosen to encompass the region of interest. The dimensions of each block were chosen so that the number of points estimated would adequately represent the surface shape. If enough grid points were not estimated, the surface would not represent the smoothness of the true surface and critical information could potentially be lost. If too many grid points were used, the computational time would be excessive. Although the grid dimensions were not required to be the same as in the structure analysis, the same grid sizes were chosen. These dimensions provided adequate density and reduced the number of graphical procedures used in displaying the results.

Using this grid structure, the kriging problem for the anthropometric data was constructed in the same manner in which kriging problems are constructed in geostatistics. Figure 3.10 illustrates the problem of estimating the midpoint of a block within the grid based on the data points in the general vicinity. Each block is partitioned by a dotted line. The objective is to determine the value $\hat{X}$ as a weighted average of the points within the range a. The value of ah was determined in the

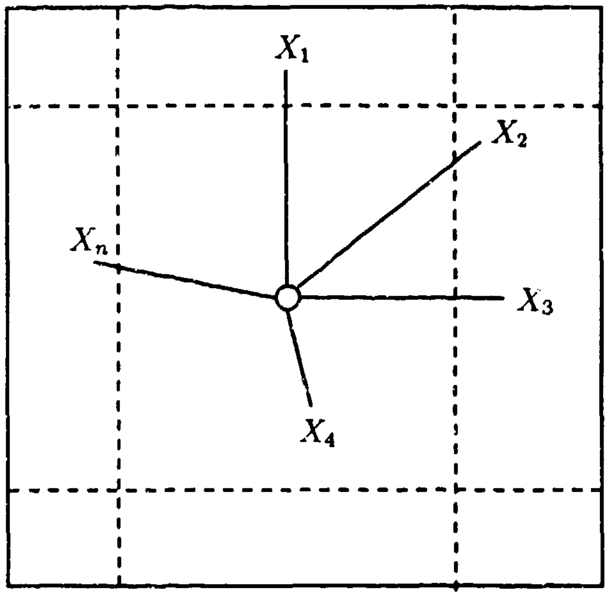

Figure 3.10. Anthropometric Kriging problem

structural analysis. The equation for this estimate is:

$$\hat{X} = w_1X_1 + w_2X_2 + \ldots + w_nX_n$$

The weights w_i are chosen to minimize the error variance

$$\sigma_\epsilon^2 = w_1\gamma(h_{1x}) + w_2\gamma(h_{2x}) + \ldots + w_n\gamma(h_{nx}) + \lambda$$

These weights are determined by solving the kriging system of equations presented in the *Literature Review*. Specifically, the universal form of kriging was used. The first step in the process was to remove the global trend from the data (reference *Structural Analysis*). To remove any residual trend, the first-order terms of a polynomial were added to the kriging matrix. Therefore, the following system of equations was solved to determine the weights:

$$\begin{bmatrix} \gamma(h_{11}) & \gamma(h_{12}) & \dots & \gamma(h_{1n}) & 1 & X_{11} & X_{21} \\ \gamma(h_{21}) & \gamma(h_{22}) & \dots & \gamma(h_{2n}) & 1 & X_{12} & X_{22} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \gamma(h_{31}) & \gamma(h_{32}) & \dots & \gamma(h_{3n}) & 1 & X_{1n} & X_{2n} \\ 1 & 1 & \dots & 1 & 0 & 0 & 0 \\ X_{11} & X_{12} & \dots & X_{1n} & 0 & 0 & 0 \\ X_{21} & X_{22} & \dots & X_{2n} & 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \\ \lambda \\ \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} \gamma(h_{1p}) \\ \gamma(h_{2p}) \\ \vdots \\ \gamma(h_{np}) \\ 1 \\ X_{1p} \\ X_{2p} \end{bmatrix}$$

To solve these equations, a program was developed in the C programming language. This program, presented in Appendix H, basically performs three tasks: removes the trend, determines the points in the zone, and estimates the mean and variance at the point. The following summarizes the programmed procedure.

The trend was calculated through the structural analysis procedure and recorded in a data file consisting of the x,y, and z coordinates of the grid midpoints. The first step of the kriging program reads in the data from the subject file, partitions this data into the grid configuration, and subtracts the trend values at the grid midpoints. The next step is to select the first point to be kriged and to determine the known points within the region. The program iteratively progresses through the ij grid blocks for which a trend value exists. In other words, the points to be kriged are the midpoints of the grid blocks in the region of the grid where data exists. Based on the structural analysis, the range for this study was a = 6.645. The program determines the known values within the range of the point ij to be kriged by considering the points in the blocks within A = 7 blocks (i.e. $i \pm 7$, $j \pm 7$). The neighborhood is ,therefore, defined by the blocks within 7 units of the block being estimated. Using the pointer structures of the C language, the values within a block were recorded in the data entry routine to enhance the efficiency of the program. Additionally, if the number of points within the neighborhood exceeded 200, the value of A was incrementally reduced by 1 thereby decreasing the neighborhood

until the number of points was between 0 and 200. The number of points in the neighborhood was limited to less than 200 to reduce the computational time and to minimize the computational errors resulting from the process of solving the system of kriging equations. After obtaining the known points in the vicinity, the program establishes the matrix form of the kriging equations. This step includes calculating the distances and variograms for each pair of points. To solve kriging equations, the LUDCMP and LUBKSB routines from *Numerical Recipes in C* were used (19, 28–45). The FORTRAN versions of these routines were discussed in the *Theoretical Variogram* section. After determining the set of weights which minimized the estimation variances, the estimates and the associated variances were recorded in output data files. This estimation process was repeated for every ij grid block of interest.

In some cases, numerical difficulties occured in kriging some of the ij blocks. More research is needed to determine the root cause of the problem. For this study, an additional kriging program was written to determine the values of the midpoints where difficulties were encountered. This additional program used the values of the midpoints in the region around the point of interest and performed the same kriging operations as discussed previously. This form of kriging was introduced as block kriging in Chapter II. The program is presented in Appendix H.

The output of the kriging programs was a file consisting of residuals and a file consisting of variances. To obtain the kriged surfaces, the trend previously removed was added to the residuals. Figure 3.11 illustrates the kriged surface for Subject 09. The kriged surfaces, residuals, and variances for all 30 subjects are located in Appendix D.

Bayesian Analysis.

After estimating each individual surface, a statistical analysis was performed to determine a recursive relationship for updating the aggregate estimates and vari-

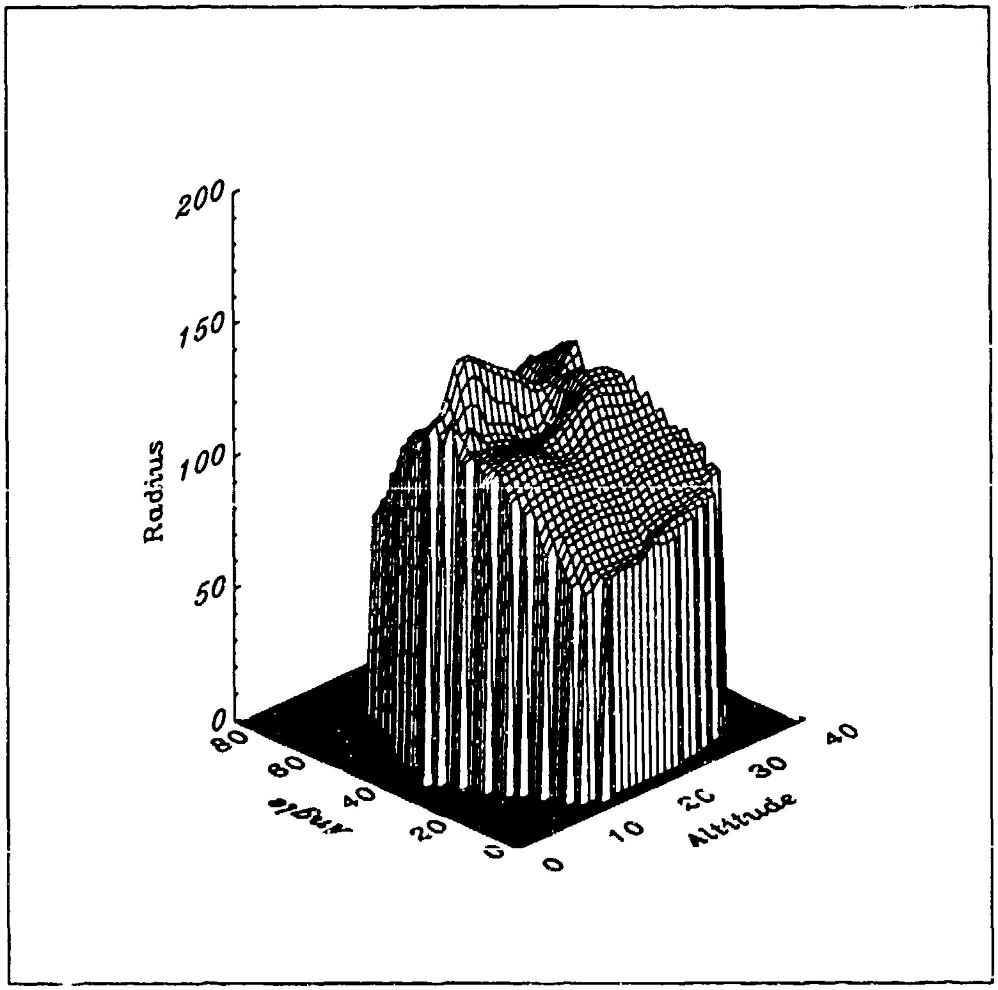

Figure 3.11. Kriged Surface for Subject 09

ances. The program in Appendix I reads in the estimates and variances for the current best estimate and for the next subject to be added in the best estimate. Using this program and the following equations, the updated value of the surface estimate ($x_{i,j}(t_n)$) and the variance of this estimate ($\sigma^2_{x_{i,j}}$) is calculated for every i, j grid point:

$$\hat{x}_{i,j}(t_n) = \hat{x}_{i,j}(t_{n-1}) + K_{i,j}(t_n)[z_{i,j} - \hat{x}(t_{n-1})]$$

and

$$\sigma^2_{x_{i,j}}(t_n) = \sigma^2_{x_{i,j}}(t_{n-1}) - K_{i,j}(t_n)\sigma^2_{x_{i,j}}(t_{n-1})$$

where

$$K_{i,j}(t_n) = \sigma^2_{\hat{x}_{i,j}(t_{n-1})}/(\sigma^2_{\hat{x}_{i,j}(t_{n-1})} + \sigma^2_{z_{i,j}})$$

and $z_{i,j}$ is the estimate for the i, j surface point of the subject being added.

As an initial starting point, the trend, determined in the structural analysis procedure, was used to estimate the surface at t_0. The initial variances at t_0 were calculated using the program in Appendix I. The mechanics of the filter permit the initial estimates of the variance to be relatively large. In this analysis, the initial variances were set equal to 20.0. The procedure was performed sequentially starting with Subject 09 and proceeding through Subject 199. Graphical representations of the surfaces are provided in Appendix D. Additionally, the surfaces estimated in the development of this procedure were the surfaces used to support the night-vision goggles study. Therefore, the surface plot of the final surface is presented in the *Surface Estimation* section.

Surface Estimation.

Using the procedures developed in the previous sections, the regions around the eyes and nose of each individual data set were kriged to determine the surfaces which minimized the error variances. The individual regions were combined using the recurvise model to provide an estimate of the surface which will support the design of the night-vision goggles. This surface is illustrated in Figure 3.12.

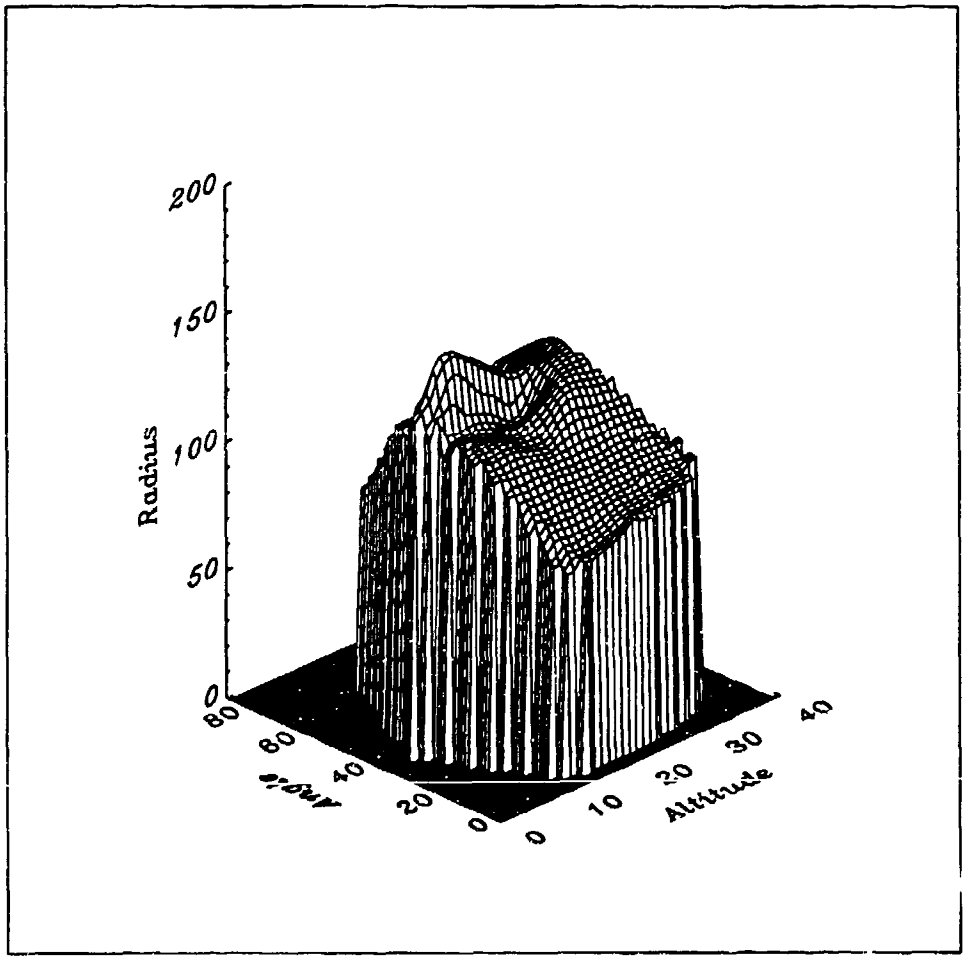

Figure 3.12. Night Vision Goggles Surface Estimate

The following summarizes the steps taken to produce the final surface estimate.

Step 1. Data Alignment. The data sets for each face were aligned so that the distances betweens the four landmarks (the Left and Right Tragions, the Glabella, and the Subnasale) and the four reference points were minimized. Furthermore, these landmarks bounded the region of interest and were logical choices for correcting the tilt in the x and y directions. The aligned data sets are illustrated in Appendix A.

Step 2. Trend Analysis. The global trend was removed to produce residual data sets. The trend is represented above in Figure 3.5. The residuals are represented in Appendix B.

Step 3. Variogram Determination. The data for the area under study was organized so that the points of each face were placed in an appropriate grid block. Based on the estimator for the experimental variogram, the variograms for the subjects were calculated and plotted. Reference Appendix C for the Variogram Plots.

Step 4. Theoretical Variogram. The theoretical variogram was determined using the weighted least-squares program and the experimental variograms calculated in *Step 3.* After analyzing the variograms of the 30 subjects used in the trend analysis, five of the subjects appeared to be outliers. Therefore, the variograms for these subjects were not included in the estimate of the theoretical variogram. This finding prompted the multivariate analysis discussed in the next section. The form and parameters of the theoretical variogram were as follows:

$$\gamma(h) = \begin{cases} 2.226(\frac{3}{2}\frac{h}{6.645} - \frac{1}{2}\frac{h^3}{6.645^3}) + 0.689 & \text{if } h < 6.645 \\ 2.226 + 0.689 & \text{if } h \geq 6.645 \\ 0 & \text{if } h = 0 \end{cases}$$

The theoretical variogram is plotted with each of the individual variograms in Appendix C. Because the theoretical variogram was determined using weighted least-squares, the plots do not reflect the emphasis of some points over others.

Step 5. Kriging of Residuals. This kriging program was designed to krige the residuals of the data sets. Using this routine, the residuals for the 25 subjects used in the variogram analysis and five additional subjects, whose variograms matched the theoretical model, were kriged to obtain the estimates and variances.

Step 6. Verfication. The results from *Step 5* were verified using the verification program. This procedure used the interpolation function of kriging to correct for numerical problems which were encountered in some instances.

Step 7. Addition of Trend. Because the trend was removed in the kriging procedure, a program to add the trend was used to determine the kriged facial region.

Step 6. Bayesian Updates. The kriged surfaces were updated using the Kalman filter developed in the previous section and the program in Appendix I for implementing this process.

The steps summarized above were developed in the previous sections and demonstrate the process used in applying kriging in the analysis of anthropomotric data. Basically, these steps provide the methodology for obtaining the surface estimates. The following section provides the details for the multivariate analysis.

Multivariate Analysis.

To investigate the feasibility of clustering the faces prior to the estimation process, a multivariate analysis of the data was performed. The intent of this analysis was to determine if groups corresponding to sizes could be identified based on several distance and angular measures. Furthermore, this analysis was performed in an attempt to determine the relationship between the faces used in the kriging analysis and the outliers identified in the structural analysis. The use of multivariate analysis methods in clustering anthropometric data is well established. Therefore, for this thesis, the multivariate analysis served only as a preliminary investigation into the

Table 3.2. Original Angular and Distance Measures

Points	Distances	Angles	Angles
0 Right Tragion	2–8	0–3–10	3–4–7
1 Right Infra Zygion	0–10	0–5–10	3–4–6
2 Right Zygofrontale	1–9	1–3–9	0–7–10
3 Glabella	3–7	1–6–9	0–6–10
4 Sellion	3–4	2–4–8	2–6–8
5 Pronasale	7–5	2–7–8	2–5–8
6 Subnasale	4–6	1–4–9	1–7–9
7 Promenton	3–6	3–4–5	
8 Left Zygofrontale	7–4		
9 Left Infra Zygion	3–5		
10 Left Tragion			

potential use of these methods in future research efforts. Specifically, this effort consisted of variable indentification, factor analysis, and cluster analysis.

Variable Indentification. The first step in the multivariate analysis was the selection of angular and distance measures which would capture the shape and size characteristics of the faces. The data for the subjects used in the structural analysis was also used for this study. Table 3.2 displays the various angle and distance measures which were used. The coordinates for the points were used to calculate the distances and the angles. Appendix K includes the programs used for developing the data files.

Factor Analysis. Using the data files created in the previous step, an analysis was performed to determine the true dimensionality of the data. The SAS factor procedure was used for the factor analysis. This procedure is described in detail in the SAS reference manual (21, 335–376). An iterative process was used to determine the final factors. This process involved the removal of variables which did not load heavily on on the factors and the rotation of the axes using the varimax rotation option to highlight the relationship between certain variables and factors. Addition-

ally, a subjective review of the factors and the contrasts was performed to determine if the resulting factors were logical and reasonable. This step produced a data file containing factor scores for the observations and provided significant insight to the underlying structure of the data.

Cluster Analysis. Given the data file of factor scores, cluster analysis was performed using the SAS cluster procedure. This procedure is explained in the SAS manual (21, 255-316). Basically, the observations were classified based on the average linkage method of the cluster procedure. The resulting groups were analyzed to determine if natural groupings or sizes were represented or if a relationship existed between the resulting groups and the outliers previously identified in the structural analysis. The computer files for the SAS routines are located in Appendix K and the results of this procedure are reported in the next chapter.

IV. Results and Conclusions

This chapter includes the results of the analysis and several conclusions based on these results. As previously stated, the purpose of this thesis was to statistically analyze anthropometric data to support improvements in the design of flight equipment. This goal was achieved. The following results and conclusions are provided with reference to the objectives outlined in Chapter I.

Results

In general, the results of this effort are the products developed to implement the procedures for analyzing anthropometric data. The procedures developed in Chapter III and the computer programs contained in the appendices provide the means for statistically analyzing the data to support improvements in the design of flight equipment. The demonstration of the procedures, using the data to support development of the night-vision goggles, produced numerical and graphical results which confirmed the hypothesis that kriging is a viable statistical procedure for estimating anthropometric surfaces. Additionally, the multivariate analysis provided experimental results which were discussed in *Facial Classification*.

Procedure Development. The first objective of this study was to develop a viable kriging procedure for estimating facial surfaces. The procedure developed for kriging anthropometric data in the preceeding chapter is a result of this research. The details included in the discussion and the programs contained in the appendices are, in essence, the physical results of this development.

Aggregation of Individual Estimates. The second objective of this thesis was to develop a recursive model for updating and aggregating the individual surface estimates. The recursive model, or Kalman filter, developed in the preceeding chapter and the programs listed in the appendices are results of this study.

Facial Surface Estimation. The third objective was to apply the kriging procedure to estimate the region of the face around the eyes and nose which will influence the design of the night-vision goggles. The surface estimate obtained using the results of the first two objectives is illustrated in Figure 3.12 in the preceeding chapter. This figure represents a surface which accounts for the shape of the facial features in the region under study and minimizes the variability between individuals. A progression of this surface, beginning with the first kriged surface, is provided in Appendix E. These representations are results of this effort.

Facial Classification. The purpose of the multivariate analysis was to determine if faces could be grouped into classes, based on various angular and distance measures, which would represent various sizes required for the flight apparatus. If sizes could be identified, the faces of a particular group could be kriged and updated independently to further reduce the variability of the surface estimates. The results of the multivariate analysis are summarized below.

Factor Analysis Results. The primary result of the factor analysis was the determination that the dimensionality of the facial data is based on five underlying factors. These factors represent five distinct features of the facial region and are illustrated in Figure 4.1. Table 4.1 provides the angular and distance measures associated with each of the factors. The first factor appears to identify width and breadth features. The second factor seems to represent the length of the faces. The third factor represents length measures within the central region of the face. Finally, the fourth and fifth factors define the protrusion of the nose with respect to the forehead and chin. A second result of the factor analysis was the data file containing the factor scores for the observations. This file was used for the classification of the faces.

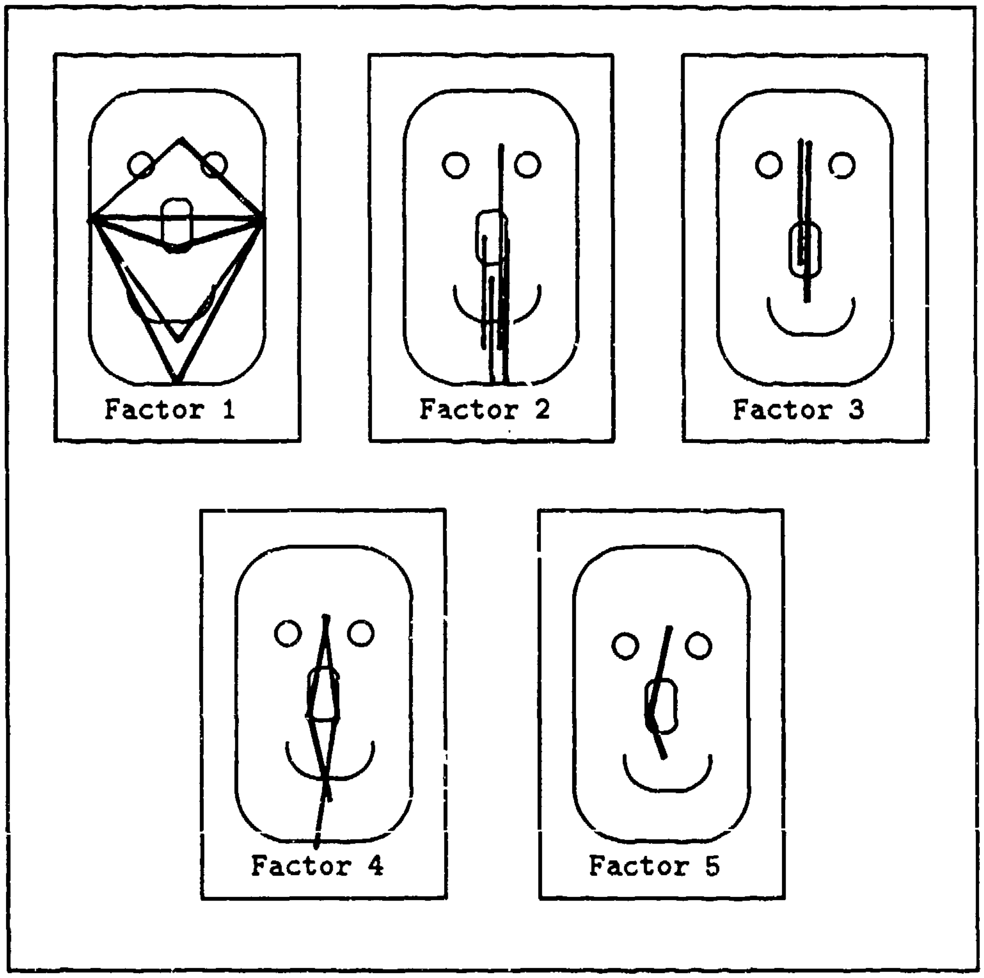

Figure 4.1. Factor Representations

Table 4.1. Factor Definitions

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Distances	1–9	4–6 3–6 7–4 3–5	3–4 3–5		
Angles	1–3–9 1–6–9 1–4–9 1–7–9			3–4–7 3–4–6	3–4–5

Table 4.2. Cluster Definitions

Clusters	Subjects
1	07*
2	01*
3	33*
4	171
5	150*
6	161
7	199 12* 14
8	151* 152 89*
9	(all remaining subjects)

Cluster Analysis Results. Based on the cluster analysis performed using the SAS procedures, the observations can be grouped into homogeneous classes based on the factor scores obtained through the factor analysis. This analysis suggests that facial classification prior to the estimation procedure merits further investigation. The clusters in Table 4.2 were identified using the average linkage method for the cluster procedure.

A major result of this analysis was the relationship between the seven faces identified as outliers in the structural analysis and the clusters. All seven outliers

(07, 01, 33, 150, 12, 151, and 89) appear in the smaller groups while the majority of the standard faces appear in the larger group. This fact suggests that the variogram analysis may also serve as a discriminating function for determining which size of an apparatus is appropriate for an individual.

Conclusions

In conclusion, this thesis develops and demonstrates the application of kriging in the statistical analysis of anthropometric data to support improvements in the design of flight equipment. Specifically, a procedure was developed for estimating the surface region in the areas where flight apparatus is worn that minimizes the variability between individuals and accounts for the shape of the region. The resulting estimates may be used by design engineers in constructing physical models to support the development of flight equipment.

In achieving the goal of this thesis, four objectives were accomplished. First, a viable kriging procedure was developed. This kriging procedure included the structural analysis of the data and the development of a universal kriging program for estimating the surfaces and the variances. Secondly, a Kalman filter was developed for updating the surface estimates. This procedure minimized the amount of storage data required to update the surfaces. Thirdly, the procedures were demonstrated in the estimation of the facial region affecting the design of the night-vision goggles. Finally, a classification of faces based on various angular and distance measurements was performed. This analysis supports the recommendation for more research in the classification of faces prior to the estimation procedure.

V. Recommendations

The primary recommendation of this thesis is the recommendation that the procedures documented in this report be used to statistically analyze anthropometric data in support of improvements in flight equipment design. Specifically, the application of kriging in estimating the surface which minimizes the variability between individual facial features and accounts for the shape of the facial region should be used in the development of physical models for flight equipment design. This process should include the alignment and structural analysis of the data sets within predetermined clusters, the actual kriging of the surfaces, and the recursive updating of the surfaces using the Kalman filter demonstrated in this study. Additionally, more research in this area is recommended.

This chapter provides recommendations which suggest either improvements in this effort or areas for further research related to this study. As this thesis may very well be the first documented application of kriging in the field of anthropometrics, further research in this area may prove promising. Recommendations are provided for all areas of this study and are presented for consideration.

Kriging

This section provides recommendations in the area of kriging.

Numerical Analysis. In some instances, numerical difficulties were encountered in obtaining the kriging estimates. The problem may be inherent to the numerical routines in the kriging program, the relatively close proximity of the data points, or some other aspect of the procedure. One recommendation is to develop an experimental procedure for determining the root cause of the computational problem. This would entail the analysis of the sample points in the regions where numerical errors occurred.

Additionally, alternative methods for matrix inversion or the solution of simultaneous equations could be considered. An initial step might be to assess the performance capability of the LUDCMP and LUBKSB routines adapted from *Numerical Recipes in C.*

Kriging Simulataneously. In this study, the surface region for each subject was kriged independently. Because of the relatively large size of the data files, this approach was logical. However, a method could be used for kriging more than one surface at a time. The potential benefits of this approach should be considered.

Determination of Grid Dimensions and Sample Sizes. As mentioned in the methodology, the grid dimensions must be chosen to support the structure of the data. Isotropic behavior, computing efficiency, and surface representation were discussed as factors to be considered. Further analysis of these factors in determining the sample and grid sizes could improve the efficiency of the procedure. Perhaps, more points or blocks were included than were necessary. A comparison of the results obtained at various dimensions may prove beneficial.

Kriging of Other Facial Regions. This study demonstrated the use of the kriging in determining the design surface for the region of the eyes and nose to support the night-vision goggles study. The techniques developed in this thesis should now be tested with other regions of the face such as the area around the mouth and nose where the oxygen masks fit.

Structural Analysis

Covariance Structure. In geostatistics, the variogram typically is used to represent the expected differences in the values of points at varying distances. However, other second-order moments, such as the covariance, may provide a better representation of the spatial relationship between points for anthropometric data. An

investigation to the use of alternative second-order structures is suggested.

Irregularly Distributed Data. The advantages of using regularly spaced grid points in the variogram calculations were discussed in the structural analysis methodology. However, the data collected with the use of the laser scanner provides an irregularly distributed configuration. A question worth investigating is what the difference is in the parameter estimates obtained by the two methods of variogram calculations.

Robustness Study. A third study concerning the structure of the variogram is recommended. This study would be to determine the robustness of the variogram structure in supporting the kriging of anthropometric surfaces. An experimental design, based on the parameters of the variogram, and a comparison of the kriging results at the experimental levels may provide a valuable assessment of the robustness of the kriging process.

Multivariate Analysis

This section provides recommendations in the area of multivariate analysis.

Initial Clustering. Two reasons were presented for clustering the data sets: to determine the relationship between the outliers in the variogram analysis with the other subjects, and to determine natural groupings which would be used to estimate sizes of the apparatus. A study in which the faces within predefined clusters are kriged independently is recommended to investigate the feasability of estimating various sizes of flight equipment.

Expanded Factor and Cluster Analysis. The data in this study was limited to the 37 subjects identified in Appendix A. Future research should include a more comprhensive data set. Additionally, a more thorough review of the variables and their relationships is suggested.

Bibliography

1. Agterberg, F. P. *Geomathematics.* New York: Elsevier Scientific Publishing Company, 1974.

2. Barnes, Randal J. and Dr. Thys B. Johnson. "Positive Kriging," *Geostatistics for Natural Resources Characterization*, Part 1, edited by G. Verly, M. David, A.G. Journel, and A. Marechal. Boston: D. Reidel Publishing Company, 1984.

3. Clark, Isobel. *Practical Geostatistics.* London: Applied Science Publishers Ltd, 1979.

4. Cressie, Noel. "Geostatistics," *The American Statistician*, Vol. 43, No. 4, 197–202 (1989).

5. David, Michael. *Geostatistical Ore Reserve Estimation.* New York: Elsevier Scientific Publishing Company, 1977.

6. Davis, John C. *Statistics and Data Analysis in Geology* (Second Edition). New York: John Wiley & Sons, 1986.

7. Dillon, William R. and Matthew Goldstein. *Multivariate Analysis: Methods and Applications.* New York: John Wiley & Sons, 1984.

8. Flury, Bernhard and Hans Riedwyl. *Multivariate Statistics: A Practical Approach.* New York: Chapman and Hall, 1988.

9. Henley, Stephen. *Nonparametric Geostatistics.* New York: John Wiley & Sons, 1981.

10. *Introduction To IDL.* Research Systems, Inc. 1989–1990.

11. *IDL User's Guide.* Research Systems, Inc. 1989.

12. Journel, A.G. "Kriging in Terms of Projections," *Mathematical Geology*, Vol. 9, No. 6, 563–586 (1977).

13. Journel, A.G. and Ch.J. Huijbregts. *Mining Geostatistics.* New York: Academic Press, 1978.

14. Matheron, G. "Principles of Geostatistics," *Statistical Analysis in Geology*, edited by John M. Cubitt and Stephen Henley. Stroudsburg: Dowden, Hutchinson & Ross, Inc., 1978.

15. Maybeck, Peter S. *Stochastic Models, Estimation, and Control*, Vol. 1. New York: Academic Press, Inc., 1979.

16. Neter, John and William Wasserman. *Applied Linear Statistical Models.* Homewood, Illinois: Richard D. Irwin, Inc., 1974.

17. Omre, Henning. "The Variogram and Its Estimation", *Geostatistics for Natural Resources Characterization*, Part 1 edited by g. Verly, M. David, A.G. Journel, and A. Marechal. Boston: D. Reidel Publishing Company, 1984.

18. Press, William H., Brian P. Flannery, Saul A. Teukolsky, and William T. Vetterling. *Numerical Recipes.* New York: Cambridge University Press, 1986.

19. Press, William H., Brian P. Flannery, Saul A. Teukolsky, and William T. Vetterling. *Numerical Recipes in C.* New York: Cambridge University Press, 1986. check date

20. Robinette, Kathleen. Research Physical Anthropologist. Personal Interview. AAMRL/HEG, Wright-Patterson AFB OH, 8 February 1990.

21. SAS Institute Inc. SAS User's Guide: Statistics, Version 5 Edition. Cary, NC: SAS Institute Inc., 1985.

22. Yakowitz, S.J. and F. Szidarovsky. "A Comparison of Kriging with Nonparametric Regression Methods," *Journal of Multivariate Analysis*, 16, 21–53 (1985).

Vita

Captain Michael Grant was born on 29 January 1960 in El Paso, Texas. He graduated from John Marshall High School in San Antonio, Texas in 1978. After high school, he attended the United States Air Force Academy where he majored in Operations Research. He completed his degree in 1982 and was assigned as a Logistics Analysis Manager for the Logistics Studies and Analysis Division of the Air Force Operational Test and Evaluation Center in Albuquerque, New Mexico. In January 1986, Captain Grant was reassigned to the Air Force Human Resources Laboratory in San Antonio, Texas where he served as Deputy Chief, Force Management Systems Branch and later served as the Executive Officer to the Commander. He earned a Master of Science degree in Industrial Engineering from St. Mary's University in San Antonio and entered the School of Engineering at the Air Force Institute of Technology in August 1988.

Captain Grant married the former Audrey Annette Valdez of San Antonio, Texas in 1982. They have two sons; Christopher Ryan and Matthew Sean.

Permanent address: 14206 Golden Woods
San Antonio, Texas 78249

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) AFIT/GOR/ENY/ENS/90M-8	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION School of Engineering	6b. OFFICE SYMBOL (If applicable) AFIT/ENY
7a. NAME OF MONITORING ORGANIZATION	
6c. ADDRESS (City, State, and ZIP Code)	7b. ADDRESS (City, State, and ZIP Code)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION AAMRL	8b. OFFICE SYMBOL (If applicable) HEG
9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER	
8c. ADDRESS (City, State, and ZIP Code) AAMRL/HEG WPAFB, OH 45433	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO / WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)
THE APPLICATION OF KRIGING IN THE STATISTICAL ANALYSIS OF ANTHROPOMETRIC DATA

12. PERSONAL AUTHOR(S)
Michael Grant, B.S., M.S., Capt, USAF

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
MS Thesis	FROM ______ TO ______		430

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Kriging, Geostatistics, Bayesian Statistics, Multivariate Analysis, Morphometrics (KR)
12	03		

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Thesis Advisors: David G. Robinson
Assistant Professor
Department of Aeronautics and Astronautics

Kenneth W. Bauer
Assistant Professor
Department of Operational Sciences

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL David G. Robinson, Asst. Professor	22b. TELEPHONE (Include Area Code) (513) 255-2362 / 22c. OFFICE SYMBOL AFIT/ENY

DD Form 1473, JUN 86 — Previous editions are obsolete. — SECURITY CLASSIFICATION OF THIS PAGE

Quality flight equipment is essential to flight crew safety and performance. Oxygen masks, night-vision goggles, and other apparatus must fit crew members comfortably and with complete functional precision. A problem currently facing the Air Force is the inconsistent quality of flight equipment. As new equipment is developed to improve crew members' performance, the requirement for design engineers to accurately account for the shape and variability of facial features becomes more critical.

This thesis develops the application of kriging in the statistical analysis of anthropometric data to support improvements in the design of flight equipment. Specifically, the geostatistical estimation technique of kriging is used to estimate the facial surfaces which influence the designs of flight apparatus. These surfaces account for the shape of the facial features and minimize the variance between individuals. A Kalman filter is developed to update and aggregate the kriged surfaces. As a proof of concept study, the techniques are demonstrated using data to support the design of the night-vision goggles currently under development. To further enhance the surface estimates, a multivariate analysis is performed to identify the factors which account for the majority of the variability between faces and to group the faces into homogenous clusters.